Author/s:
Akcay, Y;Li, Y;Natarajan, HP

Title:
Category Inventory Planning With Service Level Requirements and Dynamic Substitutions

Date:
2020-11-01

Citation:
Akcay, Y., Li, Y. & Natarajan, H. P. (2020). Category Inventory Planning With Service Level Requirements and Dynamic Substitutions. Production and Operations Management, 29 (11), pp.2553-2578. https://doi.org/10.1111/poms.13240.

Category Inventory Planning With Service Level Requirements and Dynamic Substitutions

Yalçın Akçay
Melbourne Business School, University of Melbourne, 200 Leicester Street, Carlton, VIC 3053, Australia, y.akcay@mbs.edu

Yunke Li
Department of Management, Miami Herbert Business School, 5250 University Drive, Coral Gables, FL 33124, USA yxl1601@miami.edu

Harihara Prasad Natarajan*
Department of Management, Miami Herbert Business School, 5250 University Drive, Coral Gables, FL 33124, USA hari@miami.edu

We study a single-period inventory planning problem for a category of substitutable products. This is an important practical problem facing category managers who have to maintain high service levels for constantly expanding product catalogs. We formulate the problem as a stochastic optimization model that minimizes the total stocking cost subject to service level requirements, which consist of product-specific and category-wide targets for inventory availability (ready rates) through the selling season. Our model accounts for stochastic customer arrivals, captures stockout-based substitutions, and determines initial stocking quantities jointly for all products. Recognizing the challenges that these aspects pose in solving the problem, we propose an optimization-based method that estimates the ready rates using a deterministic approximation and discretizes the selling season into a finite number of time intervals. This novel modeling approach permits us to recast the stochastic optimization model as a deterministic mixed integer linear program that can accommodate several common stockout-based substitution schemes. We characterize the worst-case behavior of this approach to develop performance guarantees. We also implemented and applied this model to randomly-generated numerical instances featuring different types of product differentiation and varying in parameter values. We observe that the approach is robust to changes in problem parameter values and yields solutions very quickly, outperforming an enumeration-based alternative, a practical heuristic, and an approach based on extant literature. Finally, we applied our approach to data from a re-seller of Information Technology products. Results illustrate that our approach scales well and has the potential to generate savings in inventory costs.

Key words: inventory planning; substitutable products; dynamic substitution; mathematical programming
History: Received: October 2018; accepted: June 2020 by Felipe Caro after two revisions.

1. Introduction

As consumer expectations for greater product variety increase, retailers have steadily expanded their catalogs to offer broader product assortments. Intense retail competition and the ease with which

* Corresponding author

consumers can resort to competitors is also forcing retailers to grow their assortments aggressively (e.g., Amazon's product selection includes over 600 million items). According to the CommerceHub Product Assortment Index, which tracks the catalogs of mass merchants, department stores, and marketplaces, traditional retailers such as Target and Walmart collectively increased their product assortments by about 80 percent in the first three quarters of 2014 alone (eCommerce Insights 2014). Naturally, this has led to an increase in inventory investments. However, an increase in inventories does not necessarily bring commensurate benefits in terms of service levels. For instance, despite growing its inventory at a faster rate than its sales, Walmart lost $3 billion in 2013 sales due to out-of-stock merchandise (Forbes 2014).

Recognizing the importance of properly matching supply and demand, and facing large out-of-stock costs and investments in inventory due to catalog expansion, firms are re-examining their inventory processes and controls to lower costs and improve service simultaneously. For *product categories*, which are defined as groups of products that consumers perceive as substitutable, determining an effective inventory policy is particularly challenging due to linkages in product demand that stem from the propensity of consumers to substitute when their primary choice is out-of-stock. These interdependencies require firms to adopt a category-wide view and address product inventory decisions jointly, incorporating how consumers choose products and how they select substitutes. Making inventory decisions individually for each product and assuming that sales are lost when a product is out-of-stock, regardless of the availability of its substitutes, could significantly and needlessly increase inventory investment. In contrast, an approach that jointly optimizes inventories for all the products in the category while satisfying service level requirements promises lower stocking costs.

Our work is motivated in part by a problem facing a large re-seller of information technology product (such as computers, software, gaming equipment) that serves markets in several countries. The company offers a vast catalog that includes products from many different vendors. Because product lifecycles are short and product costs are high in the industry, the company has been under pressure to manage inventories effectively. From an inventory perspective, the company's product portfolio is organized by categories of similar and potentially substitutable products. We consider the category manager's inventory planning problem of jointly planning the inventories of a product category to ensure high levels of service performance.

To better understand the context for this decision problem, consider the category planning process at such firms. Category planning includes both strategic and tactical considerations. Typically, firms begin their planning process by selecting products to include in each category. This strategic decision factors in consumer preferences and anticipated demand, profit margins, supplier accessibility, complementarity across categories, and the competitive landscape. The next step in the category planning process is to determine appropriate inventory levels for the chosen products. This entails judiciously selecting stocking quantities while maintaining high service levels. Consistent with the application that motivated our work, we focus on this tactical problem of ensuring appropriate inventory availability for a given selection of products. Specifically, we consider a firm that faces random demand for a multi-product category over a finite selling season. The firm begins with an initial inventory and does not have opportunities to replenish stocks during the season. Consumers arrive during the course of the season and select products based on their preferences and product availability. When their primary choice is unavailable, consumers may choose to substitute with another product. Due to this consumer-driven and dynamic out-of-stock substitution, product demands are interdependent. In turn, the stocking decisions for the products are interlinked. The category manager's goal is to minimize the total stocking cost subject to service level requirements.

The firm specifies its service level requirements using two measures. First, for each product, we consider pre-specified targets of the so-called *ready rate*, which measures the proportion of time that we expect that particular product to be available (have positive inventory) for purchase. Second, for the product category, we define and consider a *category ready rate* that measures the proportion of time that we expect a consumer would be willing to purchase any one of the products available in the category (not just their primary choice, but also possibly some other product via substitution). Our choice of availability-based measures (ready rates) rather than cycle service level requirements is based on observations in both academia and practice. Both researchers and practitioners have noted that these ready rates, which are equivalent to *fill rates* in the case of unit demand, are a better measure of inventory availability than cycle service levels as they disregard the magnitude of stockouts (Axsäter 2007). Additionally, together, the product-specific and category-based ready rate targets allow the firm to manage its service level priorities in a nuanced manner. That is, it permits the firm to potentially target a modest product ready rate, while maintaining a high category ready rate, thereby managing effectively the costs and benefits of holding inventory to meet demand. Indeed, as Kök et al. (2009) note, in contexts where consumers' inclination to substitute is high, maintaining high product-specific ready rates becomes less critical, whereas the need for a high category-level availability is accentuated. Thus, the category manager's inventory planning problem, which we refer to as the *category inventory planning* (CIP) problem, in this context entails determining initial inventories of each product in a category to minimize the total stocking cost, considering potential out-of-stock substitutions, while meeting both product and category ready rate targets.

As the literature review in Section 2 indicates, we are not aware of any work that has addressed the CIP problem. In particular, the CIP problem embeds a novel category ready rate that captures the effects of dynamic product substitution, is well-suited for practical application, and employs it as a service level requirement in inventory planning. We frame the CIP problem as a stochastic optimization model with the initial stocking quantities as decision variables, the expected product and category ready rate requirements as constraints, and the total stocking cost as the objective function. We derive expressions for the ready rates assuming that consumers arrive according to a Poisson process; however, these expectations cannot be computed in closed form. To effectively solve the CIP problem, we develop an optimization-based method that can achieve near-optimal solutions efficiently. Our optimization-based method builds on the following two key modeling ideas. First, we estimate the ready rates using a deterministic approximation of the stochastic inventory process. Next, our modeling approach discretizes the selling season into a finite number of time intervals. Together, this setup permits us to formulate and solve the CIP problem as a deterministic mixed integer linear program. This formulation is versatile and can accommodate various stockout-based substitution schemes including those that are utility-based (multinomial logit and vertical choice models) and others that are not (exogenous model). Additionally, we characterize the worst-case behavior of our approach and develop a performance guarantee by comparing bounds from our approach with those of the original problem.

One of the main goals of this paper is to develop a scalable method that can solve the CIP problem effectively in practice. Consistent with that goal, we assess the effectiveness of our model and method using both randomly-generated numerical instances (for vertical and horizontally differentiated products) and practical data (for exogenous model-based differentiation) obtained from our industry partner. The results show that our approach takes less than a minute to compute the optimal solution, on average. In contrast, a simulation-based enumeration alternative takes a few hours (even up to a few days for larger instances). The tests illustrate the robustness of our approach to changes in consumer choice schemes. Specifically, our approach is able to solve instances of vertical and horizontal product differentiation within a minute. In addition, the tests validate the effectiveness of the deterministic approximation, showing that it anticipates the stochastic inventory process closely and that it produces near-optimal solutions. We also benchmark our approach against a practical heuristic and a method adapted from existing literature. The heuristic, which assumes that demand for each product is independent, is significantly more expensive. The other benchmark is based on a static approximation of dynamic demand substitution proposed by Hopp and Xu (2008). We adapted their approach to incorporate ready rate considerations; surprisingly, it fails to consistently ensure feasibility, and even when its solutions are feasible, they are considerably more expensive than those from our approach. Finally, the computations reveal that our model can yield substantial practical benefits. When we applied our method to data from the firm, we found first that we were able to find the optimal solution even for these practical instances within a minute. Results show that the company could achieve high category ready rates while maintaining lower individual product ready rates. This flexibility in category inventory planning could lead to significant savings in inventory costs especially when the inventory investment is large.

2. Literature Review

Our work is closely related with the literature on inventory models for substitutable products. McGillivray and Silver (1978) consider a single period multi-product inventory model in which a *fixed* (*deterministic*) *proportion* of customers substitute for an out-of-stock product, and propose a heuristic for computing the optimal order quantities that minimize the total stocking cost. Focusing on the two-product case, Parlar and Goyal (1984) prove that the expected profit function is concave, and Pasternack and Drezner (1991) show that the order quantities with full substitutability can be more or less than the case with no substitutability. Building on this line of work, Rajaram and Tang (2001), and Ernst and Kouvelis (1999) show through computations that substitution can significantly impact both optimal profits and stocking levels. Netessine and Rudi (2003) generalize these earlier models to multiple substitutable products, and compare optimal inventory stocking policies in centralized and decentralized systems. Nagarajan and Rajagopalan (2008) derive the optimal inventory policy in a single period model with two partially substitutable products and negatively correlated demands, and prove the optimality of a similar policy in an infinite horizon setting. Xu et al. (2011, 2016) also consider a two-product assortment problem incorporating a flexible substitution scheme that combines product substitutions with price discounts. Bassok et al. (1999) study a single period multi-product inventory model with *full downward substitution*, which assumes that higher grade products can be used as substitutes for lower grade products. They characterize the structure of the optimal policy, and propose a gradient-based algorithm to solve the problem. Rao et al. (2004) consider a similar model to Bassok et al. (1999) with setup costs, and develop efficient algorithms that exploit the problem structure.

All of these aforementioned papers require only knowledge of the aggregate demand of each product over the entire selling season, and do not model the choice behavior of individual customers, i.e., the demand substitution is *not* dynamic. Consequently, the analysis does not entail precise information on the timing and sequence of customer arrivals. In addition, none of these papers consider service level requirements in setting the optimal stocking levels, which is an essential feature of practical planning. Our paper complements this line of literature by explicitly addressing dynamic substitutions, and introducing product-specific and category-based service level requirements to inventory management of substitutable products.

The literature on assortment planning is also pertinent to our work. The goal of assortment planning is to select the set of products and their inventory levels to maximize sales subject to budget, shelf space and various other constraints. Van Ryzin and Mahajan (1999), Cachon et al. (2005), Pan and Honhon (2012) and Davis et al. (2014) consider assortment planning problems under a variety of choice models, but only focus on the product selection aspect of the problem and ignore the stocking decision. These papers inherently assume that customers' substitution behavior is *static* and depends only on the offered assortment. In other words, customers do *not* substitute when their product choice is out of stock, essentially resulting in lost sales. This is also referred to as *assortment-based* substitution (Kök and Fisher 2007). On the other hand, Mahajan and van Ryzin (2001) study the joint inventory and assortment planning problem, by specifically considering *dynamic* substitutions that depend on the current availability of products in an assortment (*stockout-based* substitution). For a general choice model based on classical utility maximization, they show that the expected profit function is not even quasiconcave, and they propose a sample-path gradient search algorithm to determine the optimal stocking quantities. Smith and Agrawal (2000) approximate the customer demand under dynamic substitutions using the static fixed proportions substitution scheme discussed earlier. Noting that the problem leads to a discrete nonlinear optimization model with a large state space, they argue that service levels when high are "good" bounds on availability, and hence can be used to determine effective demand. Gaur and Honhon (2006) consider the same problem under a locational choice model, and derive the optimal policy structure for the static substitution case. Observing that the profit function under dynamic substitution can not be expressed in closed form, they propose the solution from the static substitution problem as a heuristic to solve the dynamic counterpart. Hopp and Xu (2008) also point to technical difficulties associated with the assortment planning problem under dynamic substitution, and present a static approximation of dynamic demand substitution based on a fluid network model. Topaloglu (2013) studies a joint stocking and product offer problem over a finite selling season under the MNL choice model and static substitutions, and formulates it as a nonlinear program with the stocking quantities and the duration of availability of each set of products as the decision variables. To control problem size, he proposes an alternative formulation dependent on the structure of the MNL model, using choice probabilities and inventory levels as decision variables. Independently, Gallego et al. (2014) also show how one can construct a linear formulation with only continuous variables when consumer choice follows the so-called General Attraction Model (a generatlization of the MNL model). Goyal et al. (2009) consider an assortment planning problem with a constraint on the total number of units that can be stocked. Customers arrive randomly with varying preferences for products, and substitute dynamically when faced with stockouts. They show that this problem is NP-hard, and develop a polynomial-time approximation algorithm for a special case. Honhon et al. (2010) study an assortment planning problem with dynamic substitutions, and adopt a demand model akin to Goyal et al. (2009). Customers are classified into types based on their preferences. Assuming that total demand is stochastic continuous and each type is a *fixed proportion* of this total demand, they model the problem as a dynamic program and present a pseudopolynomial time solution algorithm. Honhon and Seshadri (2013) show that the expected profit under the fixed proportions model in Honhon et al. (2010) is an upper bound on the expected profit of the same assortment planning problem for which each customer type is a *random proportion* of the stochastic total demand. We refer the reader to Kök et al. (2009) for a comprehensive review of assortment planning literature.

Our work differs from the assortment planning literature in several ways. First, we focus on the stocking decisions subject to service level requirements for a given assortment of products. We are not aware of any literature in this stream that plans inventories based on service-level targets. Due to its simplicity and because it does not require complex estimation procedures, practitioners prefer service-level based planning (Graves et al. 1993, Nahmias 2008); as a result, our approach is well suited for practical application. Additionally, our models employ *availability-based* measures that are favored and readily understood in industry, rather than the cycle service level measure, which researchers have noted may be deficient as a measure of service performance (Axsäter 2007). We also formulate a novel optimization model that exploits problem structure to effectively embed the key aspects of category inventory planning. The model is versatile and can readily incorporate various schemes of stockout-based substitution within the same basic framework, permitting us to solve a wide variety of instances effectively. To summarize, by developing a novel model and scalable method to an important practical problem, our work seeks to fill a gap in the literature.

3. Model Description and Formulation

Consider the single period inventory planning problem facing a firm that sells a category of n products, indexed as $j = 1, 2, \ldots, n$, over a finite selling season of length T. Let $t = 0$ and $t = T$ correspond to the beginning and the end of the season, respectively. We denote the random process of on-hand inventory over the course of selling season as $\mathbf{X} = \{\mathbf{X}(t) : 0 \leq t \leq T\}$. Here $\mathbf{X}(t) = (X_1(t), X_2(t), \ldots, X_n(t)) \in \mathbb{Z}_+^n$ is the random inventory vector at time t. The firm offers *fixed* prices $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ to customers over the selling season. Consumers arrive according to a Poisson process with rate λ and choose from the products that are available in inventory. Given an inventory level vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, we define the probability that a consumer buys product j as $\alpha_j(\mathbf{x})$. Note that the purchase probabilities $(\alpha_1(\mathbf{x}), \alpha_2(\mathbf{x}), \ldots, \alpha_n(\mathbf{x}))$ depend on inventory levels only through the set of available products, but not their quantities. Clearly, $\alpha_j(\mathbf{x}) = 0$ if $x_j = 0$. When a product is out of stock, consumers might *substitute* with another that is available; accordingly, $\alpha_j(\mathbf{x})$ incorporates all substitutions to product j. The firm's goal is to determine inventories consistent with service level targets to meet this anticipated demand. In what follows, we first introduce the details of the service level measures that the firm adopts, and then present the problem formulation.

3.1. Measuring Availability

Typically, inventory models measure a product's availability using either the probability of not stocking out (*Type–I service level*), or the fraction of demand that can be satisfied immediately from stock on hand, which is commonly referred to as the "fill rate" (*Type–II service level*). While Type–I service level is easier to compute, it does not incorporate the magnitude of shortages, and as a result provides an incomplete picture of customer service from a practical point of view (Axsäter 2007). Recognizing this limitation, we adopt the Type–II service level as our measure of availability in this paper. In the case of unit demand, Type–II service level is equivalent to the so-called *ready rate* which is defined as the fraction of time that a product has *positive* inventory. Basically, this definition reflects the *readiness* of the firm to satisfy incoming demand immediately because it has stock on hand (Axsäter 2007). We embed this logic in the ready rate service level definitions that follow.

First, we define a measure to express the availability of an individual product. Let $\omega_{jt}(\mathbf{x})$ denote the ready rate of product j given on-hand inventory vector $\mathbf{x}$ with t units of time remaining (i.e., $\mathbf{X}(T-t)=\mathbf{x}$). We express $\omega_{jt}(\mathbf{x})$ as

$$\omega_{jt}(\mathbf{x}) = \mathrm{E}\left[\frac{1}{t}\int_{T-t}^{T}\pi_j(\mathbf{X}(s))ds \,\middle|\, \mathbf{X}(T-t)=\mathbf{x}\right] \tag{1}$$

where $\pi_j(\mathbf{X}(s))$ indicates whether demand for product j can be satisfied immediately at time s. Essentially, $\pi_j(\mathbf{X}(s))$ takes the value 1 when product j's inventory is positive, i.e., $X_j(s)>0$, and 0 otherwise. We refer to $\omega_{jt}(\mathbf{x})$ as the *product ready rate*.

When customers are willing to substitute for out-of-stock products, the firm's *overall* service performance may be higher than that indicated by the product ready rates. To capture such category-wide availability, we first define *category coverage factor* as the fraction of the category's demand that can be satisfied with the stock on hand, and write it as follows

$$\Pi(\mathbf{x}) = \frac{\sum_{j=1}^{n}\alpha_j(\mathbf{x})}{\sum_{j=1}^{n}\alpha_j(\mathbf{1})}. \tag{2}$$

In (2), $\alpha_j(\mathbf{1})$ is the probability that a consumer will purchase product j when *all* products are available. Hence, $\sum_{j=1}^{n}\alpha_j(\mathbf{1})$ computes the purchase probability with *full* availability. On the other hand, $\sum_{j=1}^{n}\alpha_j(\mathbf{x})$ gives the purchase probability with available inventory $\mathbf{x}$. Therefore, (2) defines the percentage of demand that can be covered with stock on hand. Note that, $\Pi(\mathbf{x})$, in addition to $\mathbf{x}$, is also implicitly a function of the *degree* of substitutability among the products in the category. Using $\Pi(\mathbf{x})$, we can now write the *category ready rate*, denoted by $\Omega_t(\mathbf{x})$, given on-hand inventory vector $\mathbf{x}$ with t units of time remaining as

$$\Omega_t(\mathbf{x}) = \mathrm{E}\left[\frac{1}{t}\int_{T-t}^{T}\Pi(\mathbf{X}(s))ds \,\middle|\, \mathbf{X}(T-t)=\mathbf{x}\right]. \tag{3}$$

This definition of $\Omega_t(\mathbf{x})$ generalizes the concept of single product ready rate to the multi-product setting. Intuitively, we can think of $\Omega_t(\mathbf{x})$ as the fraction of time that the firm can satisfy customer demand with *at least one* of the available products. To the best of our knowledge, our paper is the first to define such an availability measure for a category of products in a dynamic substitution context. We note that other researchers have proposed assortment-based measures, but *without* incorporating stockout-based substitutions. For instance, focusing on the product selection decision in an assortment planning context, Smith and Agrawal (2000) define an assortment fill rate as "the fraction of demand that can be satisfied by items that are [selected and] normally stocked, taking [assortment-based] substitution into account, but ignoring the possibility of inventory stockouts". Their definition of the assortment fill rate is a *static* measure. In contrast, our definition in (3) reflects the dynamic nature of substitutions as the firm potentially runs out of products in the category over the course of the planning horizon.

The following two-product example illustrates how to compute $\omega_{jT}(\mathbf{x})$ and $\Omega_T(\mathbf{x})$.

Example 1. Suppose $T = 100$ and $\lambda = 0.5$. Let the purchase probabilities (which are availability dependent) be $\alpha_1(\mathbf{1}) = 0.4$, $\alpha_2(\mathbf{1}) = 0.3$, $\alpha_1(\mathbf{e}_1) = 0.6$ and $\alpha_2(\mathbf{e}_2) = 0.35$, where $\mathbf{e}_1$ and $\mathbf{e}_2$ are unit inventory vectors. For this example, given initial inventories of 6 and 10 units for product 1 and product 2, respectively, Figure 1 depicts the inventory depletion process for a *particular* sample path. During the first epoch, both products 1 and 2 are available, and inventories are depleted according to *realizations* from Poisson processes with rates $\lambda\alpha_1(\mathbf{1})$ and $\lambda\alpha_2(\mathbf{1})$, respectively. In the second epoch, only product 2 is available and is depleted with rate $\lambda\alpha_2(\mathbf{e}_2)$, which is higher than that in the first epoch, $\lambda\alpha_2(\mathbf{1})$, due to substitutions from product 1. Finally, in the last epoch the firm is out of stock. Accordingly, in the first epoch, $\pi_1(\mathbf{x}) = \pi_2(\mathbf{x}) = \Pi(\mathbf{x}) = 1$; in the second $\pi_1(\mathbf{x}) = 0, \pi_2(\mathbf{x}) = 1$, and following (2), $\Pi(\mathbf{x}) = \frac{0.35}{(0.4+0.3)} = 0.5$; and, finally $\pi_1(\mathbf{x}) = \pi_2(\mathbf{x}) = \Pi(\mathbf{x}) = 0$ in the last epoch. The *realized* product ready rates for products 1 and 2 corresponding to this particular path are $\frac{46.9}{100} = 46.9\%$ and $\frac{46.9+25.8}{100} = 72.8\%$. On the other hand, the *realized* category ready rate is $\frac{46.9+(25.8\times 0.5)}{100} = 59.8\%$. Note that to calculate $\omega_{jT}(\mathbf{x})$ and $\Omega_T(\mathbf{x})$ for a given $\mathbf{x}$, one needs to take expectations based on such computations over all sample paths of demand.

3.2. Problem Formulation

Consider the single-season planning problem facing the inventory manager of a multi-product category. Given a selection of products, the manager must determine appropriate inventory levels for each product to ensure product availability. There are two approaches that managers may consider in planning inventories – one that uses shortage/backorder costs and another based on service levels. As noted earlier, because estimating shortage costs accurately is difficult, practitioners often prefer the service level-based approach to determine inventories (Graves et al. 1993, Axsäter 2007,

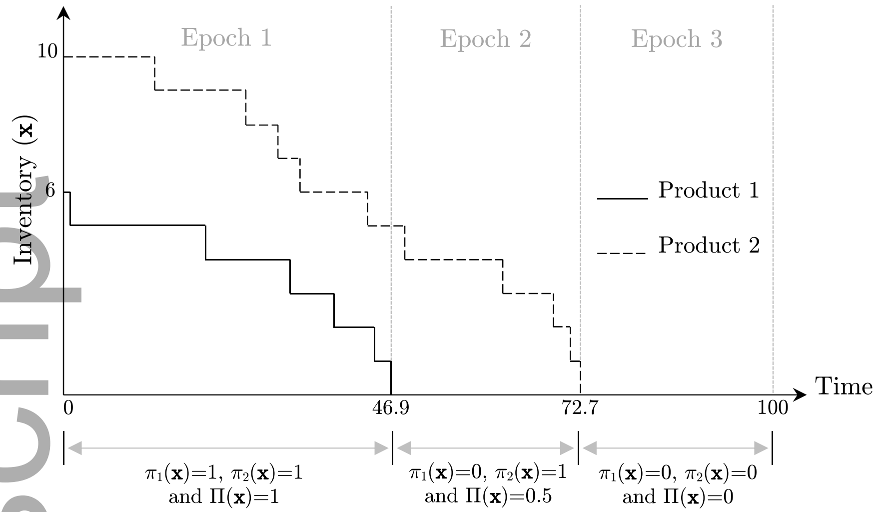


Figure 1 Sample path-based ready rate computations for a two-product example

Nahmias 2008). Along the lines of the literature on inventory planning for multi-product systems with common components (see Baker et al. 1986, Gerchak et al. 1988, Bookbinder and Tan 1988, Cohen et al. 1989, Chen and Krass 2001), we formally define the CIP problem as follows: given target product and category ready rates, the manager must determine initial inventories for each product that satisfy the ready rate requirements while minimizing the total stocking cost of the inventories.

To model the CIP problem, we first define $\mathbf{y} = (y_1, y_2, \dots, y_n)$ as the vector that denotes the inventory at the start of the selling season. Let $\mathbf{c} = (c_1, c_2, \dots, c_n)$ represent the unit purchasing cost vector. Then, the inventory manager's objective is to minimize $\mathbf{c}.\mathbf{y}$, the total inventory stocking cost. Suppose the firm sets ϕ_j as the target ready rate for each product j, and Φ as the target category ready rate. Then, we can formulate the CIP problem as the following optimization model.

$$[\text{CIP}] \qquad \min \quad \mathbf{c}.\mathbf{y} \tag{4}$$

$$s.t.$$

$$\omega_{jT}(\mathbf{y}) \geq \phi_j, \quad \text{for all } j = 1, 2, \dots, n, \tag{5}$$

$$\Omega_T(\mathbf{y}) \geq \Phi, \tag{6}$$

$$\mathbf{y} \in \mathbb{Z}_+^n \tag{7}$$

Constraint (5) requires that, over the selling season of length T, the initial inventory $\mathbf{y}$ achieve a ready rate of $\omega_{jT}(\mathbf{y})$ that meets the target ϕ_j for each product j. Similarly, constraint (6) ensures that category ready rate of inventory $\mathbf{y}$, $\Omega_T(\mathbf{y})$, meets the minimum requirement of Φ.

As we can see from (1) and (3), neither $\omega_{jT}(\mathbf{y})$ nor $\Omega_T(\mathbf{y})$ can be expressed in closed form analytically. As such, developing an appropriate method for CIP is challenging as it requires both simulation and optimization. Moreover, because both product and category ready rate requirements must be satisfied simultaneously, the CIP problem must optimize inventories *jointly* for all products. Recognizing these difficulties, we propose a novel approach to solve the CIP problem in the following section.

4. An Approximation for the CIP Problem

Our approach to modeling and solving CIP builds on a deterministic approximation of the inventory process. Many researchers (e.g., Cooper 2002) have noted that such deterministic mathematical programming approaches yield computable and reasonable policies in practice. Motivated by this observation, in this section, we build on *two* key modeling ideas to formulate the CIP problem as a deterministic mixed-integer linear program. First, rather than working with the ready rates in (1) and (3) that are not amenable to computation, we consider a deterministic flow approximation of the stochastic inventory process to estimate the ready rates. Next, building on the deterministic approximation, we propose a discretization of the selling season (i.e., we represent the season as a collection of contiguous and discrete time periods rather than a continuous span of time) to setup the model formulation. We describe these ideas next and then present our model formulation.

4.1. Approximating the Ready Rates

One of the main issues in solving the CIP problem is the difficulty in evaluating the ready rates that an inventory vector $\mathbf{y}$ can achieve. This difficulty stems from the need to evaluate all possible paths of the inventory process, which explicitly takes into account the stochastic and dynamic nature of demand, within an optimization framework. Instead, as an approximation, we propose using a *deterministic* depletion rate for the inventory, rather than a Poisson process, assuming inventory is infinitesimally divisible. With a deterministic depletion rate, the inventory process essentially follows a single path, and the corresponding ready rates are easy to compute. This approach, also known as the *fluid approximation*, has been used to approximate the evolution of stochastic processes in general and inventory processes specifically (e.g., Akçay et al. 2013, Tan and Karabatı 2013, and Honhon and Seshadri 2013).

The following algorithm provides an iterative process to compute the approximate product and category ready rates, respectively denoted by $\tilde{\omega}_{jT}(\mathbf{y})$ and $\tilde{\Omega}_T(\mathbf{y})$, for a selling season of T time units and a given initial inventory $\mathbf{y}$.

Step 0. Define $\mathbb{J}^+(\mathbf{x}) = \{j : x_j > 0\}$ for any $\mathbf{x}$.
Initialize $k = 1$, $t^k = 0$ and $\mathbf{y}^k = \mathbf{y}$.
Initialize $\tilde{\omega}_{jT}(\mathbf{y}) = 0$ and $\tilde{\Omega}_T(\mathbf{y}) = 0$.

Step 1. Calculate $d^k = \min_{j \in \mathbb{J}^+(\mathbf{y}^k)} \left\{ \frac{y_j^k}{\lambda \alpha_j(\mathbf{y}^k)}, T - t^k \right\}$.

Step 2. Update $\mathbf{y}^{k+1} = \mathbf{y}^k - \lambda \alpha_j(\mathbf{y}^k) d^k$ and $t^{k+1} = t^k + d^k$.
Update $\tilde{\omega}_{jT}(\mathbf{y}) = \tilde{\omega}_{jT}(\mathbf{y}) + \frac{d^k}{T}$ for all $j \in \mathbb{J}^+(\mathbf{y}^k)$ and $\tilde{\Omega}_T(\mathbf{y}) = \tilde{\Omega}_T(\mathbf{y}) + \frac{\Pi(\mathbf{y}^k) d^k}{T}$.
Set $k = k + 1$.

Step 3. If $\mathbb{J}^+(\mathbf{y}^k) = \varnothing$ or $t^k = T$ then Stop, else go to Step 1.

In Step 0, we define $\mathbb{J}^+(\mathbf{x})$ as the set of products with positive inventories for any given inventory vector $\mathbf{x}$. We also set the epoch index k to 1, the current time t^1 to the start of the selling season, and the first epoch's beginning inventory $\mathbf{y}^1$ to $\mathbf{y}$. Step 1 calculates the duration for which each product j is expected to last before running out, if it is depleted at a fixed rate of $\lambda \alpha_j(\mathbf{y}^k)$, and chooses the minimum of these durations as the duration of epoch k, denoted by d^k. As a result, the first stockout in $\mathbb{J}^+(\mathbf{y}^k)$ essentially marks the end of epoch k. Then, Step 2 updates the beginning inventory for the next epoch $k+1$, noting that a fixed $\lambda \alpha_j(\mathbf{y}^k) d^k$ units of product j inventory is depleted in epoch k, and progresses the current time by d^k time units to indicate the start of epoch $k+1$. Further, Step 2 also keeps record of the approximate ready rates $\tilde{\omega}_{jT}(\mathbf{y})$ and $\tilde{\Omega}_T(\mathbf{y})$ realized until this point in the selling season. Finally, Step 3 checks whether all products are stocked out or the end of the selling season is reached. If so, the algorithm stops, otherwise we return back to Step 1 for another iteration.

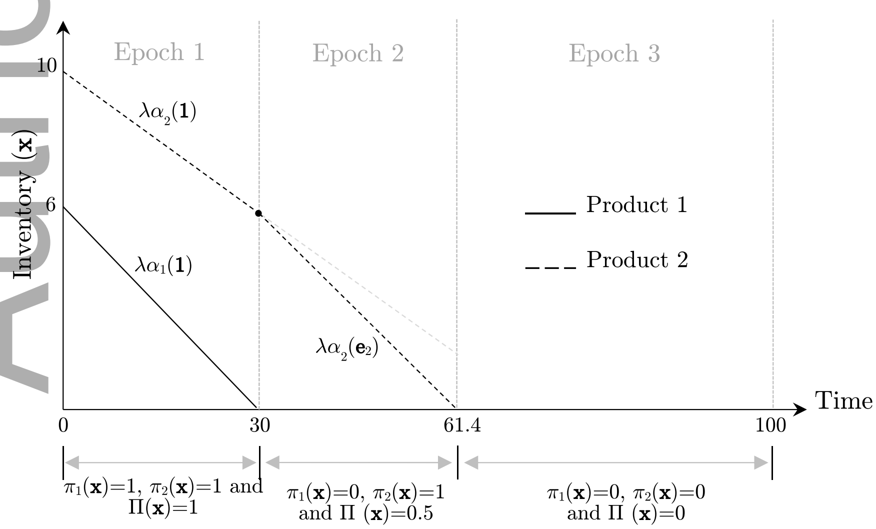


Figure 2 Approximate ready rate computations for a two-product example

For the same two-product example that we considered in Section 3.1, Figure 2 illustrates how the given initial inventories are depleted at deterministic rates (rather than based on the Poisson customer arrivals as in Figure 1). In the first iteration (epoch), both products 1 and 2 are available, and their inventories are depleted at rates $\lambda\alpha_1(\mathbf{1}) = 0.2$ and $\lambda\alpha_2(\mathbf{1}) = 0.15$, respectively. Hence, this epoch lasts $d^1 = \min\{\frac{6}{0.2}, \frac{10}{0.15}, 100\} = 30$ time units (Step 1). By the end of the first epoch, product 1 inventory is completely exhausted, whereas product 2 still has $10 - 30 \times 0.15 = 5.5$ units remaining (Step 2). At this point, we also set $\tilde{\omega}_{1T}(\mathbf{x}) = \tilde{\omega}_{2T}(\mathbf{x}) = \tilde{\Omega}_T(\mathbf{x}) = \frac{30\times 1}{100} = 30\%$ (Step 2). In the second iteration, only product 2 is available, and is depleted at rate $\lambda\alpha_2(\mathbf{e}_2) = 0.175$. Therefore, the duration of the second epoch (after the infliction point in the figure) is $d^2 = \min\{\frac{5.5}{0.175}, 100\} = 31.4$ time units (Step 1). We now update $\tilde{\omega}_{2T}(\mathbf{x}) = 30\% + \frac{31.4\times 1}{100} = 61.4\%$ and $\tilde{\Omega}_T(\mathbf{x}) = 30\% + \frac{31.4\times 0.5}{100} = 45.7\%$ (Step 2). Since neither of the two products have any stock left at the end of the second epoch, the algorithm stops (Step 3). Note that this sample path is the only path we need to consider in order to calculate the approximate ready rates. On the other hand, one can estimate the exact ready rates using Monte Carlo simulation. Let $\omega^s_{jT}(\mathbf{y})$ and $\Omega^s_T(\mathbf{y})$ denote the simulated ready rates for an initial inventory vector $\mathbf{y}$. For this example, we have $\omega^s_{1T}(\mathbf{x}) = 29.8\%$ (vs. $\tilde{\omega}_{1T}(\mathbf{x}) = 30\%$), $\omega^s_{2T}(\mathbf{x}) = 61.0\%$ (vs. $\tilde{\omega}_{2T}(\mathbf{x}) = 61.4\%$) and $\Omega^s_T(\mathbf{x}) = 45.6\%$ (vs. $\tilde{\Omega}_T(\mathbf{x}) = 45.7\%$).

4.2. Discretizing the Selling Season

The algorithm in Section 4.1 suggests a useful approximation for estimating the ready rates corresponding to a given initial inventory. Nevertheless, since the relevant computations of the approximate ready rates follow an iterative logic, it is not readily obvious if we can frame [CIP] as a mathematical program with closed-form expressions. To explore the feasibility of embedding the logic of the approximate ready rate calculations in an optimization model, we first examine the structure of these calculations. Observe that an inventory path that is based on a deterministic depletion rate, as Figure 2 illustrates, clearly specifies the time epochs, their durations, and the set of products available in each epoch. With this information, we can calculate the indicator function $\pi_i(\mathbf{x})$ and the category demand coverage factor $\Pi(\mathbf{x})$ in each epoch, and weight them with the corresponding epoch durations to determine approximate ready rates. Building on this observation, suppose we define auxiliary decision variables that (i) specify the duration of each epoch, (ii) indicate the products that are available in each epoch, and (iii) determine the purchase probability for every product in each epoch. Then, we should be able to formulate [CIP] with closed-form expressions for the approximate ready rates. However, the resulting model formulation would be nonlinear, because the ready rate computations require the multiplication of the epoch durations and the corresponding purchase probabilities (or product availability indicators), all of which would be decision variables.

To address this issue, we now propose a *discretization* scheme that permits us to formulate our problem as a linear model.

Our approach discretizes the selling season into m intervals, indexed as $k = 1, 2, \ldots, m$, with pre-specified lengths $\tau^1, \tau^2, \ldots \tau^m$ such that $\sum_{k=1}^{m} \tau^k = T$. If we assume that stockouts can occur only at the endpoints of these intervals, then, we no longer require separate decision variables for epoch durations as they are implied by the decision variables indicating availability. This permits us to express the approximate ready rates as linear functions of the decision variables. We refer to this discretized version of the CIP problem with service levels measured using deterministic approximations, and where stockouts can occur only at the endpoints of the pre-specified intervals, as the A-CIP problem (*approximate* CIP).

To formulate the A-CIP problem, consider the following decision variables:

z_j^k = 1 if product j is available in interval k; and 0 otherwise ($j = 1, 2, \ldots, n$; $k = 1, 2, \ldots, m$)
a_j^k = customer choice probability for product j during interval k ($j = 1, 2, \ldots, n$; $k = 1, 2, \ldots, m$)
y_j = initial inventory of product j ($j = 1, 2, \ldots, n$)

Then, the formulation for the A-CIP problem, which we denote as [A-CIP], follows as

$$[\text{A-CIP}] \quad \min \sum_{j=1}^{n} c_j y_j$$

$$s.t.$$

$$z_j^{k+1} \le z_j^k \qquad j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m-1 \qquad (8)$$

$$y_j \ge \sum_{k=1}^{m} \lambda \tau^k a_j^k \qquad j = 1, 2, \ldots, n \qquad (9)$$

$$\sum_{k=1}^{m} \left(z_j^k \cdot \frac{\tau^k}{T} \right) \ge \phi_j \qquad j = 1, 2, \ldots, n \qquad (10)$$

$$\sum_{k=1}^{m} \left(\frac{\sum_{j=1}^{n} a_j^k}{\sum_{j=1}^{n} \alpha_j(\mathbf{1})} \cdot \frac{\tau^k}{T} \right) \ge \Phi \qquad (11)$$

$$\left[\begin{array}{c} \text{Purchase Probability} \\ \text{Constraints} \end{array} \right] \qquad (12)$$

$$z_j^k \in \{0, 1\} \qquad j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m \qquad (13)$$

$$a_j^k \ge 0 \qquad j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m \qquad (14)$$

$$y_j \in \mathbb{Z}^+ \qquad j = 1, 2, \ldots, n. \qquad (15)$$

Constraint (8) ensures a product j that was unavailable in interval k will remain unavailable in subsequent intervals. For every product j in each interval k, given the purchase probability a_j^k, the inventory required to cover the product's demand in that interval is $\lambda \tau^k a_j^k$ (because of the deterministic rate approximation and our assumption that products stock out only at the end of intervals). Using this setup, constraint (9) makes certain that every product j has enough inventory to cover the demand that occurs during the intervals in which it is available. Observe that the purchase probability a_j^k of product j in interval k is a decision variable that must take the value of the actual purchase probability $\alpha_j([z_1^k, z_2^k, \ldots, z_j^k = 1, \ldots, z_n^k]) > 0$ if product j is available in interval k and must be zero otherwise. These purchase probability values are governed by the nature of the consumer choice model and can be expressed through linear constraints in [A-CIP]. Note that the value of a_j^k can vary with k, capturing, if needed, changing consumer preferences over the course of the selling season. Because our model can accommodate various consumer choice models, we have deferred an explicit specification of constraints (12) to Section 4.3.

Constraint (10) computes the approximate ready rate for product j by expressing the total duration of its availability (i.e., $\sum_{\forall k} z_j^k \tau^k$) as a percentage of the length of the selling season, and requires it to be at least ϕ_j. Constraint (11) computes the approximate category ready rate based on the category coverage factor in each interval k, given by $\Pi([z_1^k, z_2^k, \ldots, z_n^k]) = \sum_{j=1}^n a_j^k \Big/ \sum_{j=1}^n \alpha_j(\mathbf{1})$, and forces it to be at least Φ. Finally, constraints (13)–(15) are the binary and non-negativity requirements.

While other researchers (e.g., Gallego et al. 2004, Topaloglu (2013), Gallego et al. 2014) have proposed linear programs without resorting to binary variables for related problems in assortment planning, their modeling approaches are not applicable to the CIP problem. The Choice-Based Linear Programming approach (CBLP) approach (see Gallego et al. 2004) specifies continuous variables to define the *duration* that the firm offers a subset from the assortment (choice set), taking advantage of the notion that multiple choice sets can be offered during the selling season and that choice sets at various points of time can be independent of each other. In contrast, the CIP problem requires tracking product availability over time, thus linking the choice sets available to customers at different points of time during the selling season. Therefore, the continuous variables in the CBLP are not sufficient to model the CIP problem. The Sales-Based Linear Programming approach (SBLP) (see Topaloglu (2013) and Gallego et al. 2014), an alternative to the CBLP, defines continuous decision variables to reflect the *sales* of each product during the selling horizon. This approach does not readily extend to the CIP problem as there is no direct way to impose the ready rate constraints in [CIP] using the sales variables. Moreover, the SBLP approach is applicable only to the general attraction consumer choice model (a generalization of the multinomial logit model that we later discuss in Section 4.3.1), and cannot be used for other choice models (such as the vertical choice model in Section 4.3.1 and the exogenous model in Section 4.3.2).

4.3. Formulating the Choice Model Constraints

We next explore how to compute the purchase probabilities through a set of linear constraints, in place of (12), in [A-CIP]. There are two commonly-used schemes to determine how consumers make purchases. The first scheme is *utility-based* and consumers choose the product that maximizes their utility. Consumer utility in this case depends on the price and non-price characteristics of the available products in the category. In contrast, the second scheme does not explicitly consider consumer utilities; in this so-called *exogenous model*, consumers choose a product from the category, and if this product is unavailable, they might substitute to another product with a given probability.

In what follows, we provide sets of linear constraints that capture the salient features of each of these choice models. Note that these constraints can potentially be used in other applications that feature such consumer choice models. For expositional convenience, we assume that the underlying nature of consumer preferences and the resulting substitution behavior do not vary over time, i.e., $\alpha_j(\mathbf{x})$ is independent of t, and only changes based on the availability of products in the category.

4.3.1. Utility-Based Models. The nature of product differentiation determines how consumers derive utility from the non-price characteristics of products. A category is said to be *horizontally differentiated* when consumers have idiosyncratic preferences for the non-price characteristics of products, and is considered to be *vertically differentiated* when consumers universally agree on their product preferences (rank-ordering of products).

Horizontally differentiated category. For horizontally differentiated product categories, the multinomial logit (MNL) discrete choice model is commonly used to study the purchase behavior of consumers (McFadden 1986, Anderson et al. 1992). Let $\Psi_j = \beta_j - p_j$ be the consumers' net valuation for product j, expressed as a function of its price p_j and its valuation β_j for non-price characteristics. Then, denoting the "outside option" as product $n+1$ and normalizing its value to zero ($\Psi_{n+1} \equiv 0$) gives us the following MNL choice probabilities

$$\alpha_j(\mathbf{1}) = \begin{cases} \frac{e^{\beta_j - p_j}}{1+\sum_{i=1}^{n} e^{\beta_i - p_i}} & j = 1, 2, \dots, n \\ \frac{1}{1+\sum_{i=1}^{n} e^{\beta_i - p_i}} & j = n+1 \end{cases} \tag{16}$$

when all products in the category are available, as in the first interval k. In any subsequent interval k, the purchase probability of product j can be expressed as the following function of product availabilities:

$$a_j^k = \alpha_j([z_1^k, z_2^k, \dots, z_j^k = 1, \dots, z_n^k]) = \frac{e^{\beta_j - p_j}}{1 + \sum_{i=1}^{n} z_i^k e^{\beta_i - p_i}} \qquad j = 1, 2, \dots, n.$$

On the other hand, if product j is not available in interval k, i.e., $z_j^k = 0$, the purchase probability for product j in interval k must be zero, i.e., $a_j^k = 0$. We can also write the probability that consumers prefer the outside option in interval k as

$$a_{n+1}^k = \alpha_j([z_1^k, z_2^k, \ldots, z_n^k]) = \frac{1}{1 + \sum_{i=1}^n z_i^k e^{\beta_i - p_i}}$$

The MNL model has the so-called independence from irrelevant alternatives (IIA) property that the ratio of purchase probabilities of any two products, say j and $n+1$, is independent of the characteristics and the availability of all other products in the category. That is, the ratio

$$\rho_j = \frac{\alpha_j(\mathbf{1})}{\alpha_{n+1}(\mathbf{1})} = e^{\beta_j - p_j}, \qquad j = 1, 2, \ldots, n \tag{17}$$

is a constant parameter for product j that can be computed easily. Using (17), a_j^k simplifies to $a_j^k = \rho_j a_{n+1}^k$. With this simplification and noting that $\sum_{j=1}^{n+1} a_j^k = 1$, computing the MNL purchase probabilities reduces to solving for a_{n+1}^k. The following set of linear constraints encapsulates this logic.

$$a_j^k \le z_j^k \alpha_j(\mathbf{e}_j) \qquad j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m \tag{18a}$$

$$a_j^k \le \rho_j a_{n+1}^k \qquad j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m \tag{18b}$$

$$a_j^k \ge \rho_j a_{n+1}^k - (1 - z_j^k)\rho_j \qquad j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m \tag{18c}$$

$$a_j^k \ge \alpha_j(\mathbf{1}) z_j^k \qquad j = 1, 2, \ldots, n \tag{18d}$$

$$\sum_{j=1}^{n+1} a_j^k = 1 \qquad k = 1, 2, \ldots, m \tag{18e}$$

$$a_{n+1}^k \ge 0 \qquad k = 1, 2, \ldots, m \tag{18f}$$

In [A-CIP], for a horizontally differentiated product category, we embed (18) in place of (12) and refer to the resulting model as [A-CIP$^{\text{H}}$]. Constraints (18a), (18b) and (18c) jointly force $a_j^k = \rho_j a_{n+1}^k$ if $z_j^k = 1$, and $a_j^k = 0$ if $z_j^k = 0$, to capture the salient features of the MNL choice model. When product j is available, constraints (18d) require that a_j^k, the purchase probability of each product j, be greater than the minimum value $\alpha_j(\mathbf{1})$. Constraint (18e) makes sure that purchase probabilities in each interval k add up to 1, and constraint (18f) is the non-negativity constraint for the purchase probability of the outside option (which is the only new variable, not in [A-CIP]). The following proposition formally establishes the validity of our model. For the proof, please refer to Appendix A.1.

PROPOSITION 2. *Formulation [A-CIP$^{\text{H}}$] is valid for the A-CIP problem for a horizontally differentiated product category in the sense that an inventory vector $\mathbf{y}^F$ is feasible in [A-CIP$^{\text{H}}$] if and only if $\mathbf{y}^F$ is also a feasible solution to the A-CIP problem for a horizontally differentiated category.*

Vertically differentiated category. Consider a category in which products are vertically differentiated, with *quality* capturing all non-price characteristics. Let product j have quality q_j (common to all consumers) and assume that products can be ordered as $q_1 > q_2 > \cdots > q_n > 0$. We express a typical consumer's net utility from the purchase of product j at price p_j as $\theta q_j - p_j$, where θ represents the consumer's sensitivity to quality. In order to capture heterogeneity in consumer sensitivities, suppose θ follows a cumulative probability distribution $F(\cdot)$ with support $[0,\infty)$. Such choice models have been widely used to describe vertical demand (e.g. Tirole 1988, Wauthy 1996, Bhargava and Choudhary 2008, Akçay et al. 2010). Denoting the "outside option" as product $n+1$ and normalizing its value to zero ($q_{n+1} \equiv p_{n+1} \equiv 0$), the choice probabilities for the vertically differentiated product category can be written as:

$$\alpha_j(\mathbf{1}) = \begin{cases} 1 - F\left(\frac{p_1 - p_2}{q_1 - q_2}\right), & j = 1, \\ F\left(\frac{p_{j-1} - p_j}{q_{j-1} - q_j}\right) - F\left(\frac{p_j - p_{j+1}}{q_j - q_{j+1}}\right), & j = 2, \ldots, n, \\ 1 - \sum_{i=1}^{n} \alpha_i(\mathbf{1}) = F\left(\frac{p_n}{q_n}\right), & j = n+1, \end{cases} \tag{19}$$

when all the products are available as in the first interval k. Note the choice model in (19) implicity assumes that $\{p_j\}$ must form an increasing convex mapping of quality values $\{q_j\}$ in order to have non-zero choice probabilities for all products in the category when products are fully available (see Akçay et al. 2010 for a formal derivation of this condition). In any subsequent interval k, if product j is not available, i.e., $z_j^k = 0$, the purchase probability for product j must be zero, i.e., $a_j^k = 0$. We say that product i is *adjacent* to product j if $x_i > 0$, $x_j > 0$ and $x_r = 0$ for $i < r < j$. Suppose product ℓ is available in interval k, i.e., $z_\ell^k = 1$, and is adjacent to products $i < \ell$ and $j > \ell$, i.e., $z_i^k = z_j^k = 1$. Then, product ℓ's purchase probability as per the vertical choice model is

$$a_\ell^k = \alpha_\ell([z_1^k, \ldots, z_i^k = 1, 0, \ldots, 0, z_\ell^k = 1, 0, \ldots, 0, z_j^k = 1, \ldots, z_n^k]) = F\left(\frac{p_i - p_\ell}{q_i - q_\ell}\right) - F\left(\frac{p_\ell - p_j}{q_\ell - q_j}\right) \tag{20}$$

From (20), it is clear that a product's purchase probability depends only on the characteristics of its lower and higher quality adjacent products. Accordingly, we define the following additional decision variables:

$u_{i,j}^k =$ 1 if product j's higher quality adjacent in interval k is product i; 0 otherwise
$v_{j,i}^k =$ 1 if product j's lower quality adjacent in interval k is product i; 0 otherwise

When all products are available, the purchase probability of product j is at its minimum, $\alpha_j(\mathbf{1})$. Product j's purchase probability increases as its neighbors stock out and substitutions to product j occur. To conveniently track these changes as per (20), we define the following additional parameters. Let $\gamma_{i,j}$ denote the increase in product j's purchase probability, relative to $\alpha_j(\mathbf{1})$, when product

$i < j-1$ is its higher quality adjacent. For instances where product j itself is the highest quality product available, we define a dummy product 0 (with $q_0 = \infty > q_1$) as j's higher quality adjacent. For any product $j > 1$, we can express $\gamma_{i,j}$ as

$$\gamma_{i,j} = \begin{cases} 1 - F\left(\frac{p_{j-1}-p_j}{q_{j-1}-q_j}\right) & \text{for } i = 0, \\ F\left(\frac{p_i - p_j}{q_i - q_j}\right) - F\left(\frac{p_{j-1}-p_j}{q_{j-1}-q_j}\right) & \text{for } i = 1, 2, \ldots, j-2. \end{cases} \tag{21}$$

Analogously, let $\zeta_{j,i}$ denote the increase in product j's purchase probability, relative to $\alpha_j(\mathbf{1})$, when product $i > j+1$ is its lower quality adjacent. Then, for $j < n$, we can write $\zeta_{j,i}$ as follows

$$\zeta_{j,i} = F\left(\frac{p_j - p_{j+1}}{q_j - q_{j+1}}\right) - F\left(\frac{p_j - p_i}{q_j - q_i}\right) \qquad \text{for } i = j+2, j+3, \ldots, n+1. \tag{22}$$

Suppose product ℓ has products i and j as its lower and higher quality adjacents in interval k. Then, using (21) and (22), we can write

$$a_\ell^k = \alpha_\ell([z_1^k, z_2^k, \ldots, z_i^k = 1, 0, \ldots, 0, z_\ell^k = 1, 0, \ldots, 0, z_j^k = 1, \ldots, z_n^k]) = \alpha_j(\mathbf{1}) + \gamma_{i,\ell} + \zeta_{\ell,j} \tag{23}$$

The following set of linear constraints uses the construction in (23) to compute the purchase probabilities.

$$\sum_{i=j+1}^{n+1} v_{j,i}^k = z_j^k \qquad j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m \tag{24a}$$

$$\sum_{i=0}^{j-1} u_{i,j}^k = z_j^k \qquad j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m \tag{24b}$$

$$v_{j,i}^k \le z_i^k \qquad 1 \le j < i \le n;\ k = 1, 2, \ldots, m \tag{24c}$$

$$u_{i,j}^k \le z_i^k \qquad 0 \le i < j \le n;\ k = 1, 2, \ldots, m \tag{24d}$$

$$z_\ell^k \le 1 - v_{j,i}^k \qquad 1 \le j < \ell < i \le n+1;\ k = 1, 2, \ldots, m \tag{24e}$$

$$z_\ell^k \le 1 - u_{i,j}^k \qquad 0 \le i < \ell < j \le n;\ k = 1, 2, \ldots, m \tag{24f}$$

$$a_1^k = \alpha_1(\mathbf{1}) z_1^k + \sum_{i=3}^{n+1} \zeta_{1,i} v_{1,i}^k \qquad k = 1, 2, \ldots, m \tag{24g}$$

$$a_j^k = \alpha_j(\mathbf{1}) z_j^k + \sum_{i=0}^{j-2} \gamma_{i,j} u_{i,j}^k + \sum_{i=j+2}^{n+1} \zeta_{j,i} v_{j,i}^k \qquad j = 2, 3, \ldots, n-1;\ k = 1, 2, \ldots, m \tag{24h}$$

$$a_n^k = \alpha_n(\mathbf{1}) z_n^k + \sum_{i=0}^{n-2} \gamma_{i,n} u_{i,n}^k \qquad k = 1, 2, \ldots, m \tag{24i}$$

$$u_{i,j}^k \in \{0, 1\} \qquad 0 \le i < j \le n;\ k = 1, 2, \ldots, m \tag{24j}$$

$$v_{j,i}^k \in \{0, 1\} \qquad 1 \le j < i \le n+1;\ k = 1, 2, \ldots, m \tag{24k}$$

In the [A-CIP], for a vertically differentiated product category, we replace (12) with (24) and refer to the resulting formulation as [A-CIP$^{\text{V}}$]. If a product j is available, i.e., $z_j^k = 1$, constraints (24a) and (24b) assign its lower and higher quality adjacent products, respectively. Constraints (24c) and (24d) ensure that product j is assigned as an adjacent (lower or higher quality) only when it is available. When products i and j are adjacent, the next set of constraints, (24e) and (24f), require the adjacency condition: $x_\ell = 0$ for all $i < \ell < j$. Constraints (24g), (24h) and (24i) construct the purchase probabilities for a product considering its adjacents following (23). Finally, constraints (24j) and (24k) are the binary requirements on the additional decision variables, $u^k_{i,j}$ and $v^k_{j,i}$, not in [A-CIP].

We note that the forcing constraints (24e) and (24f), which restrict the availability of products based on the adjacent lower and higher quality products, are not required. From (21) and (22), for any product j, we can show that $\gamma_{i,j}$ is decreasing in i and $\zeta_{j,i}$ is increasing in i; therefore, the optimal solution would choose the higher quality adjacent with the largest index (nearest to j), and the lower quality adjacent with the smallest index (again, nearest to j), implying that (24e) and (24f) will always be satisfied in an optimal solution. We formally state, without proof, the validity of [A-CIP$^{\text{V}}$] in the following proposition and note that it follows a proof similar to that of the horizontal model.

Proposition 3. *Formulation* [A-CIP$^{\text{V}}$] *is valid for the* A-CIP *problem for a vertically differentiated product category in the sense that an inventory vector* $\mathbf{y}^F$ *is feasible in* [A-CIP$^{\text{V}}$] *if and only if* $\mathbf{y}^F$ *is also a feasible solution to the* A-CIP *problem for a vertically differentiated category.*

4.3.2. Exogenous Models. In the assortment optimization literature, several researchers (e.g., Anupindi et al. 1998, Smith and Agrawal 2000, Netessine and Rudi 2003, Kök and Fisher 2007, Bernales et al. 2017) consider a two-step dynamic substitution model in which consumers only make a single substitution attempt (transition) – a consumer substitutes to her second-choice with a certain *exogenous* probability if her first-choice is unavailable, and does *not* purchase from the firm if neither are available. Essentially, this model assumes that each consumer has at most two products in her preference list, but does *not* make any specific assumption regarding the underlying scheme (e.g., utility) to explain how consumers behave. Blanchet et al. (2016) propose a Markov chain-based choice model to generalize the two-step substitution model to allow the consumers to continue with substitution attempts until they find an available product or leave without a purchase. Recognizing that consumers are much less patient in practice, particularly in product recommendation contexts, Nip et al. (2017) introduce a Markov chain choice model that limits the number of transitions to a single attempt. Moreover, Kök (2003) also argues that limiting the number of substitution attempts is not too restrictive from a theoretical perspective either – they show that a multi-attempt model can effectively be approximated with a single-attempt substitution model as long as the substitution probabilities are not too high.

The exogenous model can be characterized as follows. Let μ_i be the probability that a consumer's first-choice is product i, and if product i is not available, let $\delta_{i,j}$ be the probability that a consumer chooses product j as a substitute. Note that neither μ_i nor $\delta_{i,j}$ are dependent on inventory availability.

If product j is available in interval k, i.e., $z_j^k = 1$, then

$$a_j^k = \alpha_j([z_1^k, z_2^k, \ldots, z_j^k = 1, \ldots, z_n^k]) = \mu_j + \sum_{i=1, i\neq j}^{n} \mu_i \delta_{i,j}(1 - z_i^k) \tag{25}$$

Otherwise, $\alpha_j([z_1^k, z_2^k, \ldots, z_j^k = 0, \ldots, z_n^k]) = 0$. From (25), we see that $\alpha_j(\mathbf{e}_j) = \mu_j + \sum_{i=1, i\neq j}^{n} \mu_i \delta_{i,j}$ and $\alpha_j(\mathbf{1}) = \mu_j$, and $\alpha_j(\mathbf{1})$ and $\alpha_j(\mathbf{e}_j)$ are lower and upper bounds, respectively, on a_j^k. The following linear constraints establish these relations for [A-CIP].

$$a_j^k \leq z_j^k \alpha_j(\mathbf{e}_j) \qquad j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m \tag{26a}$$

$$a_j^k \geq z_j^k \alpha_j(\mathbf{e}_j) - \sum_{i=1, i\neq j}^{n} \mu_i \delta_{i,j} z_i^k \qquad j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m \tag{26b}$$

If $z_j^k = 0$, constraint (26a) is binding; conversely, (26b) has no slack if $z_j^k = 1$ because of constraint (9) and the non-negativity of the objective function coefficients. Thus, in each case constraints (26) compute the purchase probabilities from the exogenous demand model exactly. Constraints (26) correspond to the purchase probability constraints (12) in [A-CIP] for product categories with exogenous model-based product differentiation. We refer to the [A-CIP] model with constraints (26) in place of (12) as the [A-CIP$^{\text{E}}$] formulation. The following proposition formally establishes the validity of this formulation without proof as it follows a similar logic to that of the horizontal model.

PROPOSITION 4. *Formulation* [A-CIP$^{\text{E}}$] *is valid for the* A-CIP *problem for a category whose demand follows the exogenous model in the sense that an inventory vector* $\mathbf{y}^F$ *is feasible in* [A-CIP$^{\text{E}}$] *if and only if* $\mathbf{y}^F$ *is also a feasible solution to the* A-CIP *problem for a category with the exogenous substitution scheme.*

5. Performance Bounds

By adopting a deterministic approximation of the inventory process and by discretizing the planning horizon, [A-CIP] formulates the CIP problem as a linear, mixed-integer optimization problem. The formulation provides a robust framework that is well-suited to accommodate various types of product differentiation schemes. Our optimization-based method for the CIP problem entails solving the formulation – [A-CIP$^{\text{H}}$], [A-CIP$^{\text{V}}$], or [A-CIP$^{\text{E}}$] – corresponding to the nature of product differentiation in the category.

In this section, we examine the worst-case performance of the [A-CIP] relative to the [CIP] by comparing the objective function values of the two formulations. Since it is difficult to characterize the exact objective function values of these constrained optimization problems, we derive lower and upper bounds on the optimal costs of [CIP] and [A-CIP] formulations. Using these bounds, we then establish performance guarantees for our approach. Let $\mathbf{y}^*$ denote the optimal solution to [CIP]. The following proposition provides lower and upper bounds on $\mathbf{c}.\mathbf{y}^*$, the optimal objective function value of [CIP].

Proposition 5. *Let*

$$\mathrm{LB}_{[\text{CIP}]} := \sum_{j=1}^{n} \lambda T c_j \alpha_j(\mathbf{1})\phi_j \text{ and}$$
$$\mathrm{UB}_{[\text{CIP}]} := \sum_{j=1}^{n} \left\{ c_j \max\left[\left\lceil \frac{\lambda T \alpha_j(\mathbf{e}_j)\Phi \sum_{i=1}^{n} \alpha_i(\mathbf{1})}{n\alpha_j(\mathbf{1})} \right\rceil, \lceil \lambda T \alpha_j(\mathbf{e}_j)\phi_j \rceil \right] \right\}.$$

$\mathrm{LB}_{[\text{CIP}]}$ *and* $\mathrm{UB}_{[\text{CIP}]}$ *are lower and upper bounds on the optimal objective function value of* [CIP], *respectively, i.e.,* $\mathrm{LB}_{[\text{CIP}]} \leq \mathbf{c}.\mathbf{y}^* \leq \mathrm{UB}_{[\text{CIP}]}$.

The proof of Proposition 5 (in Appendix A.2) derives $\mathrm{LB}_{[\text{CIP}]}$ by considering the minimum inventory required to satisfy the ready rate target for each product. Clearly, this is a lower bound on $\mathbf{y}^*$ as we ignore the category ready rate constraint. For the upper bound $\mathrm{UB}_{[\text{CIP}]}$, we construct two alternative feasible solutions by considering more aggressive inventory depletion scenarios than those in [CIP] (thus ensuring feasibility).

Next, we develop bounds on [A-CIP]. Let $\tilde{\mathbf{y}}$ denote the optimal solution to [A-CIP]. The following proposition provides lower and upper bounds on $\mathbf{c}.\tilde{\mathbf{y}}$, the optimal objective function value of [A-CIP].

Proposition 6. *Let*

$$\mathrm{LB}_{[\text{A-CIP}]} := \max\left\{ \lambda T \min_{1\leq j\leq n} \alpha_j(\mathbf{1}) \sum_{i=1}^{n} \phi_i c_i,\ \lambda T \Phi \min_{1\leq j\leq n} \{c_j\} \sum_{i=1}^{n} \alpha_i(\mathbf{1}) \right\}$$
$$\mathrm{UB}_{[\text{A-CIP}]} := \sum_{j=1}^{n} c_j \lceil \lambda T \alpha_j(\mathbf{1}) \rceil$$

$\mathrm{LB}_{[\text{A-CIP}]}$ *and* $\mathrm{UB}_{[\text{A-CIP}]}$ *are lower and upper bounds on the optimal objective function value of* [A-CIP], *respectively, i.e.,* $\mathrm{LB}_{[\text{A-CIP}]} \leq \mathbf{c}.\tilde{\mathbf{y}} \leq \mathrm{UB}_{[\text{A-CIP}]}$.

In the proof of Proposition 6 (in Appendix A.3), we construct the lower bound $\mathrm{LB}_{[\text{A-CIP}]}$ by calculating minimum number of periods for which each product must be available so to satisfy the ready rate constraints. As for the upper bound $\mathrm{UB}_{[\text{A-CIP}]}$, we consider a feasible solution which ensures that all products in the category are available during the entire selling season.

The following theorem bounds the ratio of the objective function value of [A-CIP] relative to that of [CIP] to assess the effectiveness of our modelling approach. It follows directly from the lower and upper bounds in Propositions 5 and 6.

THEOREM 7. $\frac{\text{LB}_{[\text{A-CIP}]}}{\text{UB}_{[\text{CIP}]}} \leq \frac{\mathbf{c}.\tilde{\mathbf{y}}}{\mathbf{c}.\mathbf{y}^*} \leq \frac{\text{UB}_{[\text{A-CIP}]}}{\text{LB}_{[\text{CIP}]}}$

To illustrate the performance guarantee in Theorem 7, we consider a 3-product category, and calculate the bounds in Propositions 5 and 6. Specifically, let $n = 3$, $\Phi = 95\%$, $\phi_j = 0.9 \times \Phi$ for $j = 1, 2, 3$, $\lambda = 150$, and $\alpha_j(\mathbf{1}) = 0.33$ for $j = 1, 2, 3$. First, suppose that the category is *vertically* differentiated. We generate the product prices and quality levels following the scheme described in Appendix B (assuming $\zeta = 0$) as $\mathbf{c} = (225, 150, 100)$, $\mathbf{q} = (2.25, 1.5, 1)$ and $\mathbf{p} = (0.6825, 0.18, 0.01)$. Using the choice model (19), we then calculate $\alpha_1(\mathbf{e}_1) = 0.697$, $\alpha_2(\mathbf{e}_2) = 0.88$, $\alpha_3(\mathbf{e}_3) = 0.99$. For this particular problem scenario, we obtain $36\% \leq \frac{\mathbf{c}.\tilde{\mathbf{y}}}{\mathbf{c}.\mathbf{y}^*} \leq 117\%$ as the performance guarantee in Theorem 7. Next, consider a horizontally differentiated product category. We assume $\mathbf{c} = (100, 100, 100)$. Using (17), we then calculate $\alpha_j(\mathbf{e}_j) = 0.97$ for $j = 1, 2, 3$. For this problem scenario, we evaluate $34\% \leq \frac{\mathbf{c}.\tilde{\mathbf{y}}}{\mathbf{c}.\mathbf{y}^*} \leq 117\%$ as the performance guarantee in Theorem 7. By providing suitable theoretical guarantees on the worst-case performance of [A-CIP], Theorem 7 establishes the value of our modeling approach. Moreover, the computational tests that we report in Section 7 show that these bounds on the worst-case performance of [A-CIP] are quite pessimistic.

6. Solution Approach

One of the challenges of using a deterministic approximation such as [A-CIP] is that its optimal solution might not be feasible for the original [CIP] model. Specifically, as in the example in Section 4.1, for a given vector $\mathbf{y}$, there may be minor differences between the simulated ready rates, $\omega^s_{jT}(\mathbf{y})$ and $\Omega^s_T(\mathbf{y})$, versus the ready rates approximated in [A-CIP]. Then, the [A-CIP] optimal solution $\tilde{\mathbf{y}}$ could potentially violate constraints (5) or (6) in [CIP]. To address this issue and ensure feasibility, we propose a greedy procedure that selectively and judiciously adds inventory to $\tilde{\mathbf{y}}$. We describe this procedure next.

6.1. Greedy Procedure for Feasibility

This procedure searches the neighborhood of the incumbent inventory vector greedily to find the most promising direction, i.e., a candidate product to add inventory, through a finite differences-type of approach (see Hoffman and Frankel 2018). Next, we describe this procedure, which we refer to as the *greedy feasibility* procedure.

STEP 0. Initialize $k = 0$.

Initialize $\mathbf{y} = \lceil \tilde{\mathbf{y}} \rceil$.

STEP 1. Set $k = k + 1$.

STEP 2. Evaluate $\Omega_T^s(\mathbf{y})$ and $\omega_{jT}^s(\mathbf{y})$ for all j using Monte Carlo simulation

STEP 3. If $\omega_{jT}^s(\mathbf{y}) \geq \phi_j$ for all j, go to STEP 4.

Otherwise, for all j such that $\omega_{jT}^s(\mathbf{y}) < \phi_j$, set $y_j = y_j + 1$. Go to STEP 1.

STEP 4. If $\Omega_T^s(\mathbf{y}) \geq \Phi$, then STOP.

Otherwise,

Evaluate $\Omega_T^s(\mathbf{y} + \mathbf{e}_j)$ for all j using Monte Carlo simulation.

Let $\ell = \arg\min_{j=1,\ldots,n} \left\{ \frac{\Omega_T^s(\mathbf{y}+\mathbf{e}_j) - \Omega_T^s(\mathbf{y})}{c_j} \right\}$ and set $\mathbf{y} = \mathbf{y} + \mathbf{e}_\ell$. Go to STEP 1.

6.2. Composite Method

Our solution approach combines the different modeling enhancements (deterministic approximation, discretization) noted in Section 4 with the methodological development (greedy procedure for feasibility) in Section 6.1 to create an effective *composite method* (CM) for solving the CIP problem. Given an instance of the CIP problem, the composite method first solves the version of the [A-CIP] that is appropriate for how consumers choose (e.g., vertical, horizontal, or exogenous) products in that category. Then, if needed, the composite method initiates the greedy procedure for feasibility.

To test our solution approach, we implemented the composite method in the C programming language with the optimization models [A-CIP$^{\text{H}}$], [A-CIP$^{\text{V}}$], and [A-CIP$^{\text{E}}$] using the CPLEX 12.5 callable optimization library.

7. Numerical Results

We conducted extensive numerical experiments to gauge the effectiveness of our model and method. First, we generated random problem instances to establish the effectiveness and efficiency of our approach in solving the CIP problem, and to assess the robustness of its performance across various problem scenarios. Next, we used data from our industry partner to show the effectiveness of our approach in a practical problem and to illustrate the potential benefits of adopting a category-based approach to inventory planning. All of the numerical results reported in this section are based on tests run on a Windows server with dual 2.4 GHz Xeon 8 core processors.

7.1. Results from Randomly Generated Problem Instances

Consistent with the goals of the computational tests, we developed a procedure to generate random CIP problem instances that varied across several key parameters. Specifically, we generated instances that varied in category size (number of products n, demand magnitude λ), product market share structure (products sharing the market equally versus not), service level requirements (product ready rate $\phi_j = \phi$ and category ready rate Φ), and nature of product differentiation (vertical versus horizontal – we study the exogenous model with data from our industry partner in Section 7.2). Table 1 provides a list of these parameters and the different values we consider for each parameter. Our numerical tests consider all the combinations of parameter values. We refer to each combination of parameter values as a problem scenario, and the specific scenarios with parameter values that are highlighted in bold in the table as our *baseline scenarios*. By systematically varying the parameters across all their possible values in Table 1, we derived a total of 432 different problem scenarios capturing a wide range of operational settings. We then generated 10 random instances for each scenario yielding a total of 4,320 instances. Appendix B describes the details of our problem generation procedure.

Parameter	**Values**
Category size: n	$\{2, \mathbf{3}, 4, 5\}$
Demand magnitude: λ	$\{25n, \boldsymbol{50n}, 100n\}$
Choice model	{VERTICAL, HORIZONTAL}
Category ready rate: Φ	$\{90\%, \mathbf{95\%}, 99\%\}$
Product ready rate: ϕ	$\{0.85 \times \Phi, \mathbf{0.90 \times \Phi}, 0.95 \times \Phi\}$
Market shares: $\alpha_j(\mathbf{1})$	$\left\{\boldsymbol{\frac{1}{n}}, \frac{2j}{n(n+1)}\right\}$

Table 1 **Key parameters for generating CIP problem scenarios**

7.1.1. Effectiveness of the Deterministic Approximation. The first step of the composite method entails solving the [A-CIP]. Without the greedy feasibility procedure, recall that the deterministic approximation embedded in [A-CIP] may fall short of the ready rate targets. Naturally, this begs the question: How effective is the deterministic approximation to begin with? What is the magnitude of the shortfall? In this first set of numerical tests, we examine the solutions to [A-CIP] in terms of their effectiveness in achieving the desired category and product ready rates.

To describe the numerical results, we define the following additional notation. For a given problem instance h, we define $\mathbf{y}^h$ as the optimal initial inventory vector given by the corresponding [A-CIP]. Further, let Φ^h and $\phi_j^h = \phi^h$, $j = 1, \ldots, n$, be the target ready rates for the category and the products, respectively, for instance h. Using Monte Carlo simulation, we estimate the *actual* ready rates achieved by $\mathbf{y}^h$, along the lines of the discussion in Section 3.1, based on 1,000 random sample paths of the problem instance. Let $\Omega_T^s(\mathbf{y}^h)$ denote the resulting category ready rate, and $\omega_{jT}^s(\mathbf{y}^h)$, $j = 1, \ldots, n$, the product ready rates.

We next define $\Delta\Phi$ and $\Delta\phi$, the category and product ready rate shortfalls, respectively, for a specific problem instance h as

$$\Delta\Phi := 100\% \times \left\{ \frac{(\Phi^h - \Omega^s_T(\mathbf{y}^h))^+}{\Phi^h} \right\} \text{ and } \Delta\phi := 100\% \times \left\{ \frac{1}{n} \sum_{j=1}^{n} \frac{(\phi^h_j - \omega^s_{jT}(\mathbf{y}^h))^+}{\phi^h_j} \right\},$$

where $x^+ := \max(0, x)$. Then, we calculate the following summary statistics:

`Avg`$(\cdot)$ – *average* of its argument over all problem instances under consideration

`Max`$(\cdot)$ – *maximum* of its argument over all problem instances under consideration

Table 2 presents these summary statistics for various problem scenarios generated by modifying a single key parameter of the baseline scenarios for the vertically and horizontally differentiated product categories, as well as for all 4,320 problem instances (as *Aggregate*). In the table, `Avg` captures the *magnitude* of the shortfall due to the deterministic approximation, whereas `Max` reflects the *range* of this shortfall. For the reader's convenience, we highlighted the results for the baseline scenarios in bold.

		HORIZONTAL				VERTICAL			
		`Avg`		`Max`		`Avg`		`Max`	
Parameters		$\Delta\Phi$	$\Delta\phi$	$\Delta\Phi$	$\Delta\phi$	$\Delta\Phi$	$\Delta\phi$	$\Delta\Phi$	$\Delta\phi$
n	2	1.4%	0.1%	2.0%	0.4%	1.4%	0.0%	1.7%	0.4%
	3	**1.0%**	**0.8%**	**1.3%**	**1.3%**	**0.6%**	**1.3%**	**1.2%**	**1.9%**
	4	0.8%	0.8%	1.0%	2.2%	0.7%	2.5%	0.9%	2.8%
	5	0.6%	1.8%	0.9%	3.1%	0.4%	3.5%	0.9%	4.1%
Φ	90%	0.2%	0.3%	0.5%	0.8%	0.2%	1.0%	0.5%	1.3%
	95%	**1.0%**	**0.8%**	**1.3%**	**1.3%**	**0.6%**	**1.3%**	**1.2%**	**1.9%**
	99%	2.4%	0.8%	2.7%	1.6%	2.4%	1.6%	2.6%	1.9%
ϕ	$0.85 \times \Phi$	0.9%	0.3%	1.4%	0.8%	0.8%	0.8%	1.2%	1.1%
	$\mathbf{0.90 \times \Phi}$	**1.0%**	**0.8%**	**1.3%**	**1.3%**	**0.6%**	**1.3%**	**1.2%**	**1.9%**
	$0.95 \times \Phi$	1.0%	2.0%	1.4%	2.7%	1.0%	2.6%	1.2%	3.0%
λ	$25n$	2.3%	0.9%	2.8%	2.0%	1.9%	1.8%	2.4%	2.3%
	$\mathbf{50n}$	**1.0%**	**0.8%**	**1.3%**	**1.3%**	**0.6%**	**1.3%**	**1.2%**	**1.9%**
	$100n$	0.4%	0.2%	0.6%	0.5%	0.3%	0.6%	0.6%	0.9%
$\boldsymbol{\alpha}_j(\mathbf{1})$	**equal**	**1.0%**	**0.8%**	**1.3%**	**1.3%**	**0.6%**	**1.3%**	**1.2%**	**1.9%**
	unequal	0.7%	1.2%	1.1%	1.8%	0.8%	2.0%	1.3%	2.4%
Aggregate		*1.3%*	*1.3%*	*6.6%*	*4.8%*	*1.3%*	*2.6%*	*4.9%*	*8.0%*

Table 2 Gaps between service levels attained by [A-CIP] solutions and target levels

First, we note that the values in Table 2 suggest that the shortfalls due to the deterministic approximation are small in magnitude with rather limited variability (relatively small `Max` values).

This clearly underlines the value of using the deterministic approximation. Interestingly, we observe that both in the vertical and horizontal cases, $\texttt{Avg}(\Delta\Phi)$ in Table 2 tends to decrease with the number of products. In our experimental design, the magnitude of the expected demand increases with the number of products as $\lambda \in \{25n, 50n, 100n\}$. Therefore, as one might expect, the deterministic approximation becomes more accurate with scale, which explains the decrease in $\texttt{Avg}(\Delta\Phi)$ with n. In contrast, we find that in both the vertical and horizontal cases, $\texttt{Avg}(\Delta\phi)$ increases with n. Note that although the total expected demand increases with the number of products, the expected demand for each product (as a first choice) remains the same. As such, one might not expect $\texttt{Avg}(\Delta\phi)$ to change significantly with n. However, the level of substitutions dramatically increases with n because of the increase in total demand. The [A-CIP] anticipates these substitutions to a large degree ensuring that the shortfall is small; however, $\texttt{Avg}(\Delta\phi)$ increases with n, highlighting the challenge of approximating the stochastic inventory process in a multi-product scenario.

Table 2 also shows that as Φ and ϕ_j increase, the gap between the target ready rates and the service performance of the [A-CIP] solution increases in majority of the problem scenarios. Intuitively, one would expect that the optimal inventory levels are increasing and convex in target service levels (due to the increasing "safety stocks" that are required to meet uncertain demand). Table 2 illustrates that the deterministic approximation approach does not add adequate buffers to handle this uncertainty. In Table 2, we further observe that as the expected demand increases, the [A-CIP] solution yields increasingly near-feasible product and category ready rates. In other words, the deterministic approximation to the stochastic inventory process improves with increasing expected demand. This result is particularly encouraging from the standpoint of practical application, where demand may be high. Finally, Table 2 demonstrates the robustness of the ready rates under the [A-CIP] solution with respect to product market shares, since both $\texttt{Avg}(\Delta\Phi)$ and $\texttt{Avg}(\Delta\phi)$ are relatively small when products in the category have identical market shares as well as when they are asymmetrical.

7.1.2. Effectiveness of the Composite Method. Next, we examine the effectiveness of the composite method (CM) in achieving the optimal solution to the CIP problem. Let $\tilde{\mathbf{y}}$ denote the optimal inventory vector from the CM. Recall that the CM includes the greedy feasibility procedure, which ensures that $\tilde{\mathbf{y}}$ is feasible for the original CIP problem. We assess the quality of the CM's solution ($\tilde{\mathbf{y}}$) by evaluating the gap between the optimal value of the original [CIP] model and that of the CM. To compute the optimal solution of [CIP], we use the following simulation-based enumeration method, which we refer to as SBE.

Step 0. Calculate $\underline{y}_j = argmin_{x\in\mathbb{Z}^+}(\omega^s_{jT}(\infty,\ldots,\infty,y_j=x,\infty,\ldots,\infty)\geq\phi_j)$, and

$\overline{y}_j = argmin_{x\in\mathbb{Z}^+}(\omega^s_{jT}(\underline{y}_1,\underline{y}_2,\ldots,y_j=x,\underline{y}_{j+1},\ldots,\underline{y}_n)\geq\Phi)$, for all $j\in\{1,\ldots,n\}$.

Set $\mathbb{Y}=\{\underline{y}_1,\ldots,\overline{y}_1\}\times\ldots\times\{\underline{y}_n,\ldots,\overline{y}_n\}\subseteq\mathbb{Z}^n$.

Initialize $\hat{\mathbb{Y}}\leftarrow\emptyset$.

Step 1. For all $\mathbf{y}\in\mathbb{Y}$, evaluate $\Omega^s_T(\mathbf{y})$ and $\omega^s_{jT}(\mathbf{y})$, $j\in\{1,\ldots,n\}$ using Monte Carlo simulation.

If $\Omega^s_T(\mathbf{y})\geq\Phi$ and $\omega^s_{jT}(\mathbf{y})\geq\phi_j$ for all j, then $\hat{\mathbb{Y}}\leftarrow\mathbf{y}\cup\hat{\mathbb{Y}}$.

Step 2. Let $\mathbf{y}_0=\arg\min_{\mathbf{y}\in\hat{\mathbb{Y}}}\ \mathbf{c}.\mathbf{y}$.

In Step 0, we reduce the enumeration space for $\mathbf{y}$ by identifying an easy-to-compute lower and upper bound on the initial inventory for each product. The lower bound, $\underline{y}_j$, is the *minimum* inventory needed to satisfy the ready rate requirement ϕ_j for product j, given that there will be no substitutions to this product, as all other products are abundantly available. The optimal solution might clearly need more units of product j when the category ready rate Φ is imposed. On the other hand, the upper bound, $\overline{y}_j$, is the *minimum* inventory of product j needed to satisfy the category ready rate Φ given that all other products have sufficient inventories to satisfy their respective product ready rates. Here, we are solely using product j to increase the category ready rate to its target level Φ. The optimal solution might require fewer units of product j because it could be potentially cheaper to increase inventory of other products in the category. Using these bounds, we form the n-dimensional space $\mathbb{Y}$ to enumerate candidates for the optimal solution. Then, in Step 1, we simulate the product and category ready rates $\Omega^s_T(\mathbf{y})$ and $\omega^s_{jT}(\mathbf{y})$ for every $\mathbf{y}\in\mathbb{Y}$, and include those solutions that satisfy the constraints of [cip] in a feasible set denoted by $\hat{\mathbb{Y}}$. Finally, Step 2 chooses the solution from $\hat{\mathbb{Y}}$ with the lowest total cost, as the optimal inventory vector $\mathbf{y}_0$.

To assess the quality of the CM's solution relative to the optimal solution, estimated by SBE, for a given problem instance h, we calculate

$$\Delta^{\text{CM}}_{\text{SBE}} = 100\% \times \left\{\frac{\mathbf{c}^h.\tilde{\mathbf{y}}^h - \mathbf{c}^h.\mathbf{y}^h_0}{\mathbf{c}^h.\mathbf{y}^h_0}\right\}.$$

In Table 3, for the baseline scenarios and varying n, we report the average optimality gaps `Avg`$(\Delta^{\text{CM}}_{\text{SBE}})$, maximum gaps `Max`$(\Delta^{\text{CM}}_{\text{SBE}})$, the problem size for the corresponding [a-cip] (in number of variables and constraints), as well as the CPU times for CM and SBE. First, note that the SBE approach takes an enormous amount of time to solve, especially for larger product categories. Despite modestly low simulation times for the ready rate evaluations, SBE takes time because of the large number of enumerative steps involved. Indeed, when the number of products increases to 4 and 5, the SBE approach did not converge to a solution even after 10 days of computation for all the vertically differentiated problem instances and all but two of the horizontally differentiated

		CM			SBE		
	n	Variables	Constraints	CPU (min)	CPU (min)	Avg($\Delta^{\text{CM}}_{\text{SBE}}$)	Max($\Delta^{\text{CM}}_{\text{SBE}}$)
VERTICAL	2	10002	13003	<1	16	0.07%	0.69%
	3	18003	22004	<1	685	0.49%	1.51%
	4	28004	33005	<1			
	5	40005	46006	<1			
HORIZONTAL	2	5002	11003	<1	10	0.00%	0.00%
	3	7003	16004	<1	256	0.42%	1.39%
	4	9004	21005	<1	13,773*	0.00%*	0.00%*
	5	11005	26006	<1			

*: Calculated over 2 problem instances

Table 3 **Problem size and computational performance of the composite method for the baseline scenarios**

instances. In contrast, the composite method takes less than a minute for all problem instances, even though it also requires simulations within its greedy feasibility procedure. For the smaller categories ($n=2$ and $n=3$), the SBE approach did converge, albeit after a long time, permitting us to evaluate the quality of the composite method's solution. In these cases, the optimality gap is extremely small ($< 0.5\%$, on average, with the maximum gap $< 2\%$), indicating that the composite method produces near-optimal solutions.

As Table 3 shows clearly the size of the [A-CIP] model (both in terms of number of variables and constraints) increases with n. Recall that [A-CIP$^{\text{V}}$] requires two sets of additional variables to identify products of adjacent quality and additional constraints to link them to other decision variables in the formulation, whereas [A-CIP$^{\text{H}}$] requires just one set of additional probability variables corresponding to the outside option. As a result, [A-CIP$^{\text{V}}$] is consistently larger in size than [A-CIP$^{\text{H}}$]. Nevertheless, the [A-CIP] takes less than a minute to find the optimal solution. The greedy feasibility procedure is robust, taking about five iterations on average (not reported in the table) to ensure feasibility of the [A-CIP] solution, showing that the deterministic approximation is consistently tight.

7.1.3. Flexibility of Category-Based Planning. One of the benefits of the category-based planning approach is the ability to ensure high category ready rates while requiring lower product ready rates. To understand the flexibility that this nuanced approach offers category planners, we adopt the following benchmark. Let Z^{Φ}_{ϕ} denote the optimal cost of CM when the target category ready rate is Φ and the target product ready rates are equal to $\phi_j = \phi$ for $j = 1, \ldots, n$. Consider a category manager who is focused on individual product-based planning. To achieve a category ready rate of Φ, the manager would likely have adopted inventory levels that are consistent with $\phi = 0.99 \times \Phi$ and would have incurred a cost of $Z^{\Phi}_{0.99\times\Phi}$. Of course, category managers can achieve the same category rate of Φ at a potentially lower total cost if they were to choose $\phi < 0.99 \times \Phi$. To assess the savings that stem from this flexibility, for any Φ, we compute the corresponding *category planning flexibility* as $\frac{Z^{\Phi}_{0.99\times\Phi} - Z^{\Phi}_{\phi}}{Z^{\Phi}_{0.99\times\Phi}}$. When this ratio is high, the firm can benefit greatly from adopting a category-based view of inventory planning. We note that this category planning flexibility metric is likely to understate the savings from our approach because the benchmark of $Z^{\Phi}_{0.99\times\Phi}$ assumes that the inventory planner would take into account all the interdependencies in the demands of the products, rather than plan purely based on the intrinsic demand (without substitutions), which is perhaps more likely in practice (we investigate such an independent demand-based inventory planning heuristic in Section 7.1.4). Nevertheless, we adopt this benchmark to clearly isolate the benefits of category ready rate-based inventory planning. Table 4 reports the cost savings, as measured by the category planning flexibility ratio, for the baseline scenarios with $\Phi = 95\%$ and $\phi \in \{0.80\times\Phi, 0.85\times\Phi, 0.90\times\Phi, 0.95\times\Phi, 0.99\times\Phi\}$.

		ϕ				
	n	$0.80\times\Phi$	$0.85\times\Phi$	$0.90\times\Phi$	$0.95\times\Phi$	$0.99\times\Phi$
HORIZONTAL	2	0.0%	0.0%	0.0%	0.0%	0.0%
	3	5.5%	5.5%	5.0%	4.4%	0.0%
	4	5.4%	5.2%	5.1%	3.5%	0.0%
	5	6.1%	5.9%	4.9%	3.4%	0.0%
VERTICAL	2	6.5%	5.8%	4.7%	3.8%	0.0%
	3	10.3%	8.9%	6.7%	4.1%	0.0%
	4	12.1%	9.5%	6.9%	4.4%	0.0%
	5	13.7%	10.4%	7.3%	4.3%	0.0%

Table 4 Category planning flexibility for the baseline scenarios ($\lambda = 50n$, $\Phi = 95\%$, and $\alpha_j(\mathbf{1}) = \frac{1}{n}$ for $j = 1, \ldots, n$)

For both horizontally and vertically differentiated product categories, and all category sizes, Table 4 illustrates that given the target category ready rate of $\Phi = 95\%$, category planning flexibility increases with lower levels of individual product rates, i.e., the customers would still enjoy the same category service level while the firm realizes significant cost savings. Note that costs of all products in the horizontally differentiated categories are identical (as per our problem generation scheme). Therefore, cost saving opportunities are more limited relative to the vertically differentiated categories.

7.1.4. Benchmarks. We first consider a benchmark based on the independent demand model, which assumes that all demand for a product is lost when the product is unavailable (see Gallego et al. 2014). Honhon et al. (2010) discuss a similar model in which the customer decides the product they want from the initial assortment (assortment-based substitution) and does not substitute when their chosen product is out of stock. Accordingly, the independent demand model estimates that each product attracts independent arrivals at a rate determined by the overall customer arrival process, and a fixed probability of purchase for each product, i.e., $\lambda T\alpha_j(\mathbf{1})$. Anticipating this demand stream, the manager then determines an appropriate level of inventory based on the service level requirements. Because of its simplicity, we can think of such an approach as a suitable facsimile of practice.

Given the individual product and category ready rate requirements of the CIP problem, the natural way to generate a feasible solution is to plan each product's inventory based on the category ready rate, rather than its product ready rate. We refer to this approach as the independent demand-based (IDB) method. Formally, we describe the IDB method as follows: First, we approximate the aggregate Poisson demand for product j, for $j=1,\ldots,n$, with a Normal distribution with mean $\lambda T\alpha_j(\mathbf{1})$ and standard deviation $\sqrt{\lambda T\alpha_j(\mathbf{1})}$, and set the ready rate target for the product as Φ. We then calculate y_j, for $j=1,\ldots,n$, by solving the following equation:

$$(1-\Phi)\lambda T\alpha_j(\mathbf{1})=\sqrt{\lambda T\alpha_j(\mathbf{1})}f\left(\frac{y_j-\lambda T\alpha_j(\mathbf{1})}{\sqrt{\lambda T\alpha_j(\mathbf{1})}}\right)-(y_j-\lambda T\alpha_j(\mathbf{1}))F\left(\frac{y_j-\lambda T\alpha_j(\mathbf{1})}{\sqrt{\lambda T\alpha_j(\mathbf{1})}}\right), \quad (27)$$

where $f(\cdot)$ and $F(\cdot)$ are the pdf and cdf of the standard Normal distribution. Note that (27) is the classical equation to calculate the inventory level to achieve a desired Type-II service level (see Chopra and Meindl 2016). Finally, we set $\mathbf{y}_{\text{IDB}}=\lceil\mathbf{y}\rceil$ to satisfy the integrality constraints of the CIP problem.

By not explicitly incorporating demand dependencies, the IDB method underestimates the demand, and in turn the inventory, for each product (as it does not account for $\alpha_j(\mathbf{x})\geq\alpha_j(\mathbf{1})$). However, by planning for a ready rate of Φ for each product it, overstates the service level targets, effectively increasing the inventory requirements. To assess the impact of these countervailing effects, we applied the IDB method to our test instances. Accordingly, we calculate the incremental cost of the IDB solution relative to the solution of our composite method for a given problem instance h as follows:

$$\Delta_{\text{CM}}^{\text{IDB}}=100\%\times\left\{\frac{\mathbf{c}^h.\mathbf{y}_{\text{IDB}}^h-\mathbf{c}^h.\tilde{\mathbf{y}}^h}{\mathbf{c}^h.\tilde{\mathbf{y}}^h}\right\}.$$

Tables 5 and 6 provide evidence of potential savings that our approach might generate over current practice. In addition, comparison of CM with the IDB method as a benchmark also provides insight into the value of incorporating category information (namely, substitutions and ready rates) in inventory planning.

First, we note that the IDB method yields a feasible solution to the CIP problem in all variants of the baseline scenario. Nonetheless, IDB solutions are significantly more expensive than those

	HORIZONTAL			VERTICAL		
	ϕ			ϕ		
n	$0.85\times\Phi$	$0.90\times\Phi$	$0.95\times\Phi$	$0.85\times\Phi$	$0.90\times\Phi$	$0.95\times\Phi$
2	4.8%	4.4%	4.2%	7.5%	6.6%	5.7%
3	6.0%	5.4%	4.6%	10.4%	8.9%	5.2%
4	6.2%	6.0%	5.0%	10.7%	8.0%	4.9%
5	6.1%	5.0%	3.1%	12.1%	7.6%	5.0%

Table 5 Inventory savings of CM relative to IDB, $\texttt{Avg}(\Delta_{\text{CM}}^{\text{IDB}})$, for the baseline scenarios ($\lambda = 50n$, $\Phi = 95\%$, and $\alpha_j(\mathbf{1}) = \frac{1}{n}$ for $j = 1,\ldots,n$)

	HORIZONTAL			VERTICAL		
	Φ			Φ		
n	90%	95%	99%	90%	95%	99%
2	3.8%	4.4%	6.5%	5.6%	6.6%	10.5%
3	4.2%	5.4%	9.7%	6.1%	8.9%	15.3%
4	1.6%	6.0%	11.1%	3.0%	8.0%	17.2%
5	1.5%	5.0%	12.1%	3.9%	7.6%	17.9%

Table 6 Inventory savings of CM relative to IDB, $\texttt{Avg}(\Delta_{\text{CM}}^{\text{IDB}})$, for the baseline scenarios ($\lambda = 50n$, $\phi = 0.90\times\Phi$, and $\alpha_j(\mathbf{1}) = \frac{1}{n}$ for $j = 1,\ldots,n$)

of our approach (CM) – on average 5.8% for horizontally differentiated categories, and 8.8% for vertically differentiated categories. For both product differentiation schemes and for all category sizes, Table 5 shows that cost savings with CM increase when the inventory manager can relax individual product ready rate requirements for a given category ready rate target. In Table 6, we see that the performance of the IDB method deteriorates significantly when the category ready rate target increases, for given targets of product ready rates, uniformly across all test instances. These results illustrate that the IDB method would provide cost-effective inventory plans only when the product ready rates are set sufficiently close to the category ready rate, and the method is most costly when the firm sets a high category ready rate and is willing to trade it off with lower product ready rates.

The second benchmark we consider is based on an approximation idea in Hopp and Xu (2008). Recognizing the difficulty in analyzing dynamic substitution behavior, Hopp and Xu (2008) propose a static approximation that is tractable and amenable to analysis. They represent the stochastic inventory process by a constant service rate, assume that demand is a continuous random variable, and then approximate the dynamic demand substitution process by a fluid network with a memoryless flow property. By doing so, they complement the fluid approximation that other researchers such as Mahajan and van Ryzin (2001) had developed, with a further simplification that yields a static approximation for the case when consumer demand follows an attraction model (see, e.g., Basuroy and Nguyen 1998). As they concede, their approach does *not* generalize to other choice models, such as the vertical and exogenous choice models. Therefore, we modified their approach to yield a benchmark for the CIP problem for a horizontally differentiated category.

We refer to our adaptation of Hopp and Xu (2008) as the HX method, and describe it formally as follows: We assume that the total demand over the selling season, denoted by N, follows a Normal distribution with mean λT and standard deviation $\sqrt{\lambda T}$. Hopp and Xu (2008) approximate the availability of product j as a constant service rate s_j. Their availability measure is equivalent to the expected proportion of time that demand for product j can be met, which we define as $\omega_{jT}(\mathbf{x})$ in (1). Building on this, our adaptation of their approach incorporates the service level requirements by setting the ready rate for product j as $s_j = \Phi$, for $j = 1, \ldots, n$, to capture both the product and category ready rate requirements. We then compute the initial inventory vector that is consistent with such a service rate as per Hopp and Xu (2008), as follows

$$y_j = \frac{\Phi\alpha_j(\mathbf{1})}{\alpha_0(\mathbf{1}) + \Phi\sum_j \alpha_j(\mathbf{1})}\frac{1}{E(1/N)}. \tag{28}$$

Finally, we set $\mathbf{y}_{\text{HX}} = \lceil \mathbf{y} \rceil$ to satisfy the integrality constraints of the CIP problem.

The strength of this method relies critically on how well s_j approximates the stochastic availability process of product j. When it does, we would expect $\mathbf{y}_{\text{HX}}$ to be feasible for the CIP problem. However, the HX method does not always generate a feasible solution for the CIP problem. To establish the effectiveness of our composite method relative to the HX method, we applied the HX method to baseline scenarios that feature horizontally differentiated product categories with $\Phi \in \{90\%, 95\%, 99\%\}$ and $\phi \in \{0.85 \times \Phi, 0.90 \times \Phi, 0.95 \times \Phi\}$. If $\mathbf{y}_{\text{HX}}$ is feasible, we calculate the incremental cost of the HX solution relative to the solution of our composite method for a given problem instance h as follows:

$$\Delta_{\text{CM}}^{\text{HX}} = 100\% \times \left\{ \frac{\mathbf{c}^h.\mathbf{y}_{\text{HX}}^h - \mathbf{c}^h.\tilde{\mathbf{y}}^h}{\mathbf{c}^h.\tilde{\mathbf{y}}^h} \right\}.$$

On the other hand, if $\mathbf{y}_{\text{HX}}$ is infeasible, we calculate $\Delta\Phi_{\text{HX}}$ and $\Delta\phi_{\text{HX}}$, the category and product ready rate shortfalls, respectively, for a specific problem instance h as

$$\Delta\Phi_{\text{HX}} := 100\% \times \left\{ \frac{(\Phi^h - \Omega_T^s(\mathbf{y}_{\text{HX}}^h))^+}{\Phi^h} \right\} \text{ and } \Delta\phi_{\text{HX}} := 100\% \times \left\{ \frac{1}{n}\sum_{j=1}^{n} \frac{(\phi_j^h - \omega_{jT}^s(\mathbf{y}_{\text{HX}}^h))^+}{\phi_j^h} \right\},$$

where $\omega_{jT}^{\text{S}}(\mathbf{y}_{\text{S}}^h)$ and $\Omega_T^{\text{HX}}(\mathbf{y}_{\text{HX}}^h)$ denote the product and category ready rates achieved by the inventory vector $\mathbf{y}_{\text{HX}}^h$, based on 1,000 random sample paths of the problem instance h. Table 7 documents these results along with the percentage of problem instances for which $\mathbf{y}_{\text{HX}}$ does *not* satisfy the ready rate requirements (denoted as % *infeasible*).

n	% *infeasible*	$\Delta\Phi_{\text{HX}}$	$\Delta\phi_{\text{HX}}$	$\texttt{Avg}(\Delta^{\text{HX}}_{\text{CM}})$	$\texttt{Max}(\Delta^{\text{HX}}_{\text{CM}})$
2	42.4%	2.4%	1.8%	6.2%	11.7%
3	35.6%	1.9%	1.5%	6.5%	11.9%
4	39.2%	1.7%	1.4%	6.1%	11.7%
5	38.7%	1.4%	1.4%	5.7%	11.6%

Table 7 Performance of the HX method for the baseline scenarios ($\lambda = 50n$, choice model = HORIZONTAL, and $\alpha_j(\mathbf{1}) = \frac{1}{n}$ for $j = 1, \ldots, n$)

In Table 7, we observe that the HX method indeed fails to generate feasible solutions for a significantly large proportion (more than one third) of the problem instances, with average shortfalls above 1%. A closer look at the data reveals that majority of these infeasible instances occur when the ready rate targets are relatively high, which is the case for most practical scenarios. Even when the HX solutions are feasible, they are substantially more expensive than those generated by the CM, by an average of above 5%. We note that the product costs in the problem instances for horizontally differentiated categories are all identical. Hence, the cost savings with CM are purely based on its ability to generate solutions with lower levels of inventory, while still satisfying the ready rate requirements.

7.2. Practical Application

We applied the composite method to data from a re-seller of IT products, which include computer equipment, software, gaming, networking, printing, and peripherals. The company offers a vast product assortment of over 20,000 SKUs from almost 200 vendors in more than 40 countries. The scope of the catalogue, which was expected to continue to grow with the addition of new and popular product categories, entailed a significant burden on the company's operations. Specifically, managing inventories effectively was vital for the company to successfully contend with short product lifecycles, heightened global competition, and high inventory costs. For planning purposes, the company divided its product portfolio into multiple business divisions and within each division, products were classified into about 40 categories and further into 180 subcategories (for example, the computer category included the notebook, netbook, and ultrabook subcategories). The size of each category (or subcategory) varied greatly from just a few to tens of products. We examine the effectiveness of our method by applying it to data from a single market for two product categories – ENTRY LEVEL and MID-RANGE – that belong to the computer equipment product group. Table 8 reports the number and costs of the products in these categories.

With rapid technology development and new product adoption, the length of the selling season for these product categories is short relative to the lead times for replenishment orders. Consequently, the company had only a single ordering opportunity before the start of the selling season to meet

Product Category	# of SKUs (n)	Market Share ($\boldsymbol{\alpha}_j(\mathbf{1})$) (*Min., Avg., Max.*)	Cost ($\mathbf{c}$) (*Min., Avg., Max.*)
ENTRY LEVEL	20	(0.1%, 5%, 17%)	($400, $450, $540)
MID-RANGE	13	(1%, 8%, 16%)	($600, $640, $740)

Table 8 Summary of key application data

the anticipated demand. These characteristics of the problem context suggested that our model and method may be a good fit for this context. Moreover, the categories we considered consist of a large number of products that vary in their market shares and costs. The composite method requires the intrinsic demand (without substitutions) for each product and the proportion of product i customers that will purchase a product $j \neq i$ when i is out of stock. So, our first step was to develop a method to estimate these parameters. Many researchers have developed statistical methods to estimate these parameters from sales and inventory data (please refer to Berbeglia et al. (2018) for an extensive review of such methods in retail operations). We followed one such approach described in the literature (e.g., Anupindi et al. 1998, Kök and Fisher 2007) to estimate intrinsic demand and product substitutions that follow the specifications of the *exogenous* demand model.

The main steps of the statistical estimation procedure are as follows. The procedure begins with daily sales and inventory data for the two product categories. The procedure assumes that customer arrivals for each product follows a Poisson process with a constant rate. With this rate and product substitution probability (say $\delta_{i,j}$), we can write expressions for the likelihood of observed demand. Then, using maximum likelihood estimation, the procedure then finds the parameters λ, $\alpha_j(\mathbf{1})$, and $\delta_{i,j}$ consistent with the exogenous model. With these parameter estimates, we were able to apply data from our industry partner to the CM with the [A-CIP^{E}] model.

Although the company had specific service performance goals, we applied the data with several alternative target ready rate combinations ($\Phi \in \{90\%, 95\%, 99\%\}$ and $\phi_j = \phi \in \{0.80 \times \Phi, 0.85 \times \Phi, 0.90 \times \Phi, 0.95 \times \Phi, 0.99 \times \Phi\}$) to understand the potential cost savings for each choice of Φ. These experiments are along the lines of those reported in Section 7.1.3, where we examined the inventory savings that our category-based planning approach can generate by ensuring high levels of service at the category level while lowering individual product ready rates. As before, we calculate the category planning flexibility as $\frac{Z^{\Phi}_{0.99\times\Phi} - Z^{\Phi}_{\phi}}{Z^{\Phi}_{0.99\times\Phi}}$ and report our results in Table 9.

We observe that in each case, as one might expect, given a target category ready rate, category planning flexibility increases with lower levels of individual product rates, complementing earlier results under utility-based choice models. Table 9 also shows that the firm has category planning flexibility, albeit modest, with both the ENTRY LEVEL and the MID-RANGE categories. To put this in context, we first note that even small percentage improvements in costs can be significant when the

		Φ					
		ENTRY LEVEL			MID-RANGE		
		99%	95%	90%	99%	95%	90%
ϕ	$0.99 \times \Phi$	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	$0.95 \times \Phi$	0.38%	0.44%	0.49%	0.27%	0.14%	0.14%
	$0.90 \times \Phi$	0.92%	0.91%	1.03%	0.17%	0.27%	0.27%
	$0.85 \times \Phi$	1.40%	1.43%	1.38%	0.79%	0.47%	0.47%
	$0.80 \times \Phi$	2.02%	2.01%	2.1%	1.1%	0.53%	0.53%

Table 9 Category planning flexibility for ENTRY LEVEL and MID-RANGE product categories

inventory investment is large. Additionally, recall that our estimate of the savings is conservative and the actual savings for practical instances is likely to be higher. Nevertheless, the savings illustrate that category planning flexibility is likely to be highest when the level of product substitutability in the category is relatively high, and when the lower cost products in the category are substitutes for the higher cost products. Thus, the potential savings from category-based planning depend vitally on consumer choice and the nature of product differentiation.

8. Conclusion

Managing the competing considerations of ensuring high availability and maintaining low costs is a considerable challenge for inventory managers facing uncertain demand. For categories of products, this inventory planning problem is especially challenging because of the interdependencies in product demand due to customer substitution, and the need to incorporate availability-based service level targets for both individual products and the overall category. We present a structured approach to formulate and solve this problem. Our novel modeling approach exploits the structure of the problem to formulate the category inventory planning problem as a mixed-integer linear program. Our model incorporates the complex interdependencies between product demands and the tradeoffs between availability and cost in a systematic manner. By proposing a new, versatile model-driven approach for availability-based category inventory planning this paper fills a gap in the literature.

Our approach is also well-suited for retail industrial application, particularly in B2C contexts where behaviour of *individual* consumers need to be taken into account explicitly. As evidenced by the application of our method to data from the IT product re-seller, it can solve practical instances of the problem effectively. In addition, our approach is scalable because it does not resort to simulation to solve large instances. Finally, the model and method are versatile, able to incorporate various types of substitution schemes, and different context-specific restrictions and considerations. Thus, by developing a scalable and robust approach for an important practical problem, this work can serve as an important framework for category inventory management in practice. Our model and method can also apply to a single product stocking decision in a distribution network context. In this case, one can think of the collection of distribution centers as the "category", with each distribution center requiring an individual ready rate and the network, as a whole, requiring a category ready rate.

Acknowledgements

The authors would like to acknowledge the review team for their insightful and constructive feedback that enhanced the paper. Harihara P. Natarajan would also like to acknowledge the support of the Center for International Business Education and Research that supported this work with a travel grant.

References

Akçay Y, Boyacı T, Zhang D (2013) Selling with money-back guarantees: The impact on prices, quantities, and retail profitability. *Production and Operations Management* 22(4):777–791.

Akçay Y, Natarajan H, Xu SH (2010) Joint dynamic pricing of multiple perishable products under consumer choice. *Management Science* 56(8):1345–1361.

Anderson S, DePalma A, Thisse J (1992) *Discrete choice theory of product differentiation* (MIT press).

Anupindi R, Dada M, Gupta S (1998) Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science* 17(4):406–423.

Axsäter S (2007) *Inventory control*, volume 90 (Springer Science & Business Media).

Baker KR, Magazine MJ, Nuttle HLW (1986) The effect of commonality on safety stock in a simple inventory model. *Management Science* 32(8):982–988.

Bassok Y, Anupindi R, Akella R (1999) Single-period multiproduct inventory models with substitution. *Operations Research* 47(4):632–642.

Basuroy S, Nguyen D (1998) Multinomial logit market share models: Equilibrium characteristics and strategic implications. *Management science* 44(10):1396–1408.

Berbeglia G, Garassino A, Vulcano G (2018) A comparative empirical study of discrete choice models in retail operations, *Working Paper*, Available at SSRN: https://ssrn.com/abstract=3136816.

Bernales PJ, Guan Y, Natarajan HP, Gimenez PS, Tajes MXA (2017) Less is more: Harnessing product substitution information to rationalize skus at intcomex. *Interfaces* 47(3):230–243.

Bhargava H, Choudhary V (2008) Research note-when is versioning optimal for information goods? *Management Science* 54(5):1029–1035.

Blanchet J, Gallego G, Goyal V (2016) A markov chain approximation to choice modeling. *Operations Research* 64(4):886–905.

Bookbinder J, Tan J (1988) Strategies for the probabilistic lot-sizing problem with service-level constraints. *Management Science* 34(9):1096–1108.

Cachon GP, Terwiesch C, Xu Y (2005) Retail assortment planning in the presence of consumer search. *Manufacturing & Service Operations Management* 7(4):330–346.

Chen FY, Krass D (2001) Inventory models with minimal service level constraints. *European Journal of Operational Research* 134(1):120–140.

Chopra S, Meindl P (2016) *Supply Chain Management: Strategy, Planning, and Operation* (Pearson Higher Ed), 6 edition.

Cohen MA, Kleindorfer PR, Lee HL (1989) Near-optimal service constrained stocking policies for spare parts. *Operations Research* 37(1):104–117.

Cooper WL (2002) Asymptotic behavior of an allocation policy for revenue management. *Operations Research* 50(4):720–727.

Davis JM, Gallego G, Topaloglu H (2014) Assortment optimization under variants of the nested logit model. *Operations Research* 62(2):250–273.

eCommerce Insights (2014) Product assortment index provides important lessons for retailers heading into the holidays. `http://blog.commercehub.com` Retrieved November 17, 2014.

Ernst R, Kouvelis P (1999) The effects of selling packaged goods on inventory decisions. *Management Science* 45(8):1142–1155.

Forbes (2014) How walmart could solve its inventory problem and improve earnings. `http://onforb.es/SnR9Zt` Retrieved May 22, 2014.

Gallego G, Iyengar G, Phillips R, Dubey A (2004) Managing flexible products on a network. Technical report, Columbia University Department of Industrial Engineering and Operations Research, New York.

Gallego G, Ratliff R, Shebalov S (2014) A general attraction model and sales-based linear program for network revenue management under customer choice. *Operations Research* 63(1):212–232.

Gaur V, Honhon D (2006) Assortment planning and inventory decisions under a locational choice model. *Management Science* 52(10):1528–1543.

Gerchak Y, Magazine MJ, Gamble AB (1988) Component commonality with service level requirements. *Management Science* 34(6):753–760.

Goyal V, Levi R, Segev D (2009) Near-optimal algorithms for the assortment planning problem under dynamic substitution and stochastic demand. *Submitted, June* .

Graves SC, Kan AHGR, Zipkin PH (1993) *Logistics of production and inventory*, volume 4 (Elsevier).

Hoffman JD, Frankel S (2018) *Numerical methods for engineers and scientists* (CRC press).

Honhon D, Gaur V, Seshadri S (2010) Assortment planning and inventory decisions under stockout-based substitution. *Operations Research* 58(5):1364–1379.

Honhon D, Seshadri S (2013) Fixed vs. random proportions demand models for the assortment planning problem under stockout-based substitution. *Manufacturing & Service Operations Management* 15(3):378–386.

Hopp WJ, Xu X (2008) A static approximation for dynamic demand substitution with applications in a competitive market. *Operations Research* 56(3):630–645.

Kök AG (2003) *Management of product variety in retail operations*. Ph.D. thesis, The Wharton School, University of Pennsylvania.

Kök AG, Fisher M (2007) Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research* 55(6):1001–1021.

Kök AG, Fisher ML, Vaidyanathan R (2009) Assortment planning: Review of literature and industry practice. *Retail Supply Chain Management*, 99–153 (Springer).

Mahajan S, van Ryzin G (2001) Stocking retail assortments under dynamic consumer substitution. *Operations Research* 49(3):334–351.

McFadden D (1986) The choice theory approach to market research. *Marketing Science* 5(4):275–297.

McGillivray AR, Silver EA (1978) Some concepts for inventory control under substitutable demand. *INFOR* 16(1):47–63.

Nagarajan M, Rajagopalan S (2008) Inventory models for substitutable products: optimal policies and heuristics. *Management Science* 54(8):1453–1466.

Nahmias S (2008) *Production and operations analysis* (McGraw-Hill/Irwin), sixth edition.

Netessine S, Rudi N (2003) Centralized and competitive inventory models with demand substitution. *Operations Research* 51(2):329–335.

Nip K, Wang Z, Wang Z (2017) Assortment optimization under a single transition model. *Working Paper* Available at SSRN 2916110.

Pan XA, Honhon D (2012) Assortment planning for vertically differentiated products. *Production and Operations Management* 21(2):253–275.

Parlar M, Goyal S (1984) Optimal ordering decisions for two substitutable products with stochastic demands. *Opsearch* 21(1):1–15.

Pasternack BA, Drezner Z (1991) Optimal inventory policies for substitutable commodities with stochastic demand. *Naval Research Logistics* 38(2):221–240.

Rajaram K, Tang CS (2001) The impact of product substitution on retail merchandising. *European Journal of Operational Research* 135(3):582–601.

Rao US, Swaminathan JM, Zhang J (2004) Multi-product inventory planning with downward substitution, stochastic demand and setup costs. *IIE Transactions* 36(1):59–71.

Smith SA, Agrawal N (2000) Management of multi-item retail inventory systems with demand substitution. *Operations Research* 48(1):50–64.

Tan B, KarabatıS (2013) Retail inventory management with stock-out based dynamic demand substitution. *International Journal of Production Economics* 145(1):78–87.

Tirole J (1988) *The Theory of Industrial Organization* (MIT Press).

Topaloglu H (2013) Joint stocking and product offer decisions under the multinomial logit model. *Production and Operations Management* 22(5):1182–1199.

Van Ryzin G, Mahajan S (1999) On the relationship between inventory costs and variety benefits in retail assortments. *Management Science* 45(11):1496–1509.

Wauthy X (1996) Quality choice in models of vertical differentiation. *The Journal of Industrial Economics* 44:345–353.

Xu H, Yao DD, Zheng S (2011) Optimal control of replenishment and substitution in an inventory system with nonstationary batch demand. *Production and Operations Management* 20(5):727–736.

Xu H, Yao DD, Zheng S (2016) Optimal policies for a two-product inventory system under a flexible substitution scheme. *Production and Operations Management* 25(6):1088–1105.

Appendix

A. Proofs of Analytical Results

A.1. Proof of Proposition 2

We establish this proposition by first characterizing a feasible solution (starting inventory vector) $\mathbf{y}^F$ to the A-CIP problem for a horizontally differentiated product category (A-CIP$^{\text{H}}$ problem). Then, we show that the corresponding solution $(\mathbf{y}^F, \mathbf{z}, \mathbf{a})$ is feasible in [A-CIP$^{\text{H}}$] if and only if $\mathbf{y}^F$ is also a feasible solution to the A-CIP$^{\text{H}}$ problem.

Characterizing a feasible solution of the A–CIP problem

Recall that the A-CIP problem discretizes the selling season into m intervals with pre-specified lengths $\tau^1, \tau^2, \ldots \tau^m$ and requires that stockouts occur only at the endpoints of these intervals. Let $\mathcal{T}$ be the set of such time interval endpoints, i.e., $\mathcal{T} = \{\tau^1, \tau^1 + \tau^2, \ldots, \sum_{k=1}^m \tau^k = T\}$. Further, we define $\mathcal{T}(\mathbf{y}^F) \subseteq \mathcal{T}$ as a subset of $\mathcal{T}$ indicating the times at which products stock out when a given initial inventory $\mathbf{y}^F$ is depleted according to the deterministic rates in the fluid approximation. In particular, let $\mathcal{T}(\mathbf{y}^F) = \{t_1, t_2, \ldots, t_{\eta(\mathbf{y}^F)}\}$, where $\eta(\mathbf{y}^F)$ denotes the number of products that stock out before the end of the selling season, when starting with $\mathbf{y}^F$, and t_i denotes the instant at which the i^{th} stockout occurs. We also define $\mathbb{J}^-$ as the set of products that stock out before T, and $\mathbb{J}^+$ as the set of products that do *not* stock out during the selling season. Let $j(i)$ be the index of the product corresponding to the i^{th} stockout. Accordingly, we have $\mathbb{J}^- = \{j(i) : i \leq \eta(\mathbf{y}^F)\}$. With this notation, we know that

$$y^F_{j(1)} = \lambda \alpha_{j(1)}(\mathbf{1}) t_1$$

$$\begin{aligned}
y^F_{j(2)} &= \lambda\alpha_{j(2)}(\mathbf{1})t_1 + \lambda\alpha_{j(2)}(1-\mathbf{e}_{j(1)})(t_2-t_1) \\
y^F_{j(3)} &= \lambda\alpha_{j(3)}(\mathbf{1})t_1 + \lambda\alpha_{j(3)}(1-\mathbf{e}_{j(1)})(t_2-t_1) + \lambda\alpha_{j(3)}(1-\mathbf{e}_{j(1)}-\mathbf{e}_{j(2)})(t_3-t_2) \\
&\vdots \\
y^F_{j(\eta(\mathbf{y}^F))} &= \lambda\alpha_{j(\eta(\mathbf{y}^F))}(\mathbf{1})t_1 + \lambda\alpha_{j(\eta(\mathbf{y}^F))}(1-\mathbf{e}_{j(1)})(t_2-t_1) \\
&\quad + \ldots + \lambda\alpha_{j(\eta(\mathbf{y}^F))}\Big(1-\sum_{i=1}^{\eta(\mathbf{y}^F)-1}\mathbf{e}_{j(i)}\Big)(t_{\eta(\mathbf{y}^F)} - t_{\eta(\mathbf{y}^F)-1})
\end{aligned} \tag{29}$$

Suppose $n > \eta(\mathbf{y}^F)$, i.e., not all products in the category stock out before the end of the selling season. Then, for a product j which has positive inventory after $s_{\eta(\mathbf{y}^F)}$, i.e., $j \in \mathbb{J}^+$, we can write

$$\begin{aligned}
y^F_j &= \lambda\alpha_j(\mathbf{1})t_1 + \lambda\alpha_j(1-\mathbf{e}_{j(1)})(t_2-t_1) + \ldots + \lambda\alpha_j\Big(1-\sum_{i=1}^{\eta(\mathbf{y}^F)-1}\mathbf{e}_{j(i)}\Big)(t_{\eta(\mathbf{y}^F)} - t_{\eta(\mathbf{y}^F)-1}) \\
&+ \lambda\alpha_j\Big(1-\sum_{i=1}^{\eta(\mathbf{y}^F)}\mathbf{e}_{j(i)}\Big)(T - t_{\eta(\mathbf{y}^F)})
\end{aligned} \tag{30}$$

While y^F_j can be greater than the RHS of the above equation, it will be exactly equal in an optimal solution (due to cost minimization). Let $\tilde{\omega}_{jT}(\mathbf{y}^F)$ and $\tilde{\Omega}_T(\mathbf{y}^F)$ denote the product and category service levels achieved by $\mathbf{y}^F$. Then, following the iterative logic described in Section 4.1, we know that

$$\tilde{\omega}_{j(i),T}(\mathbf{y}^F) = \frac{t_{j(i)}}{T}, \quad \forall i \le \eta(\mathbf{y}^F), \text{ and} \tag{31}$$

$$\tilde{\omega}_{jT}(\mathbf{y}^F) = 1, \quad \forall j \in \mathbb{J}^+ \tag{32}$$

Further, we have

$$\begin{aligned}
\tilde{\Omega}_T(\mathbf{y}^F) &= 1.\frac{t_1}{T} + \Pi(\mathbf{1}-\mathbf{e}_{j(1)})\frac{(t_2-t_1)}{T} + \ldots + \Pi\Big(1-\sum_{i=1}^{\eta(\mathbf{y}^F)-1}\mathbf{e}_{j(i)}\Big)\frac{(t_{\eta(\mathbf{y}^F)} - t_{\eta(\mathbf{y}^F)-1})}{T} \\
&\quad + \Pi\Big(1-\sum_{i=1}^{\eta(\mathbf{y}^F)}\mathbf{e}_{j(i)}\Big)\frac{(T - t_{\eta(\mathbf{y}^F)})}{T}
\end{aligned} \tag{33}$$

where $\Pi(\cdot)$ is the category coverage factor as defined in (2). Clearly, if $\mathbf{y}^F$ is feasible, then $\tilde{\omega}_{jT}(\mathbf{y}^F) \ge \phi_j$ for all j, and $\tilde{\Omega}_T(\mathbf{y}^F) \ge \Phi$.

To summarize, a feasible solution $\mathbf{y}^F$ to the A-CIP problem is characterized by (29) and (30), and achieves the service levels given in (31), (32) and (33).

Characterizing $(\mathbf{y}^F, \mathbf{z}, \mathbf{a})$

Starting with $\mathbf{y}^F$, we can construct a solution $(\mathbf{y}^F, \mathbf{z}, \mathbf{a})$ that is valid for [A-CIP$^{\text{H}}$], and achieves the same total stocking cost $\mathbf{c}.\mathbf{y}^F$, and service performance, as expressed in (31), (32) and (33). Let κ_i denote the index of the time interval at the end of which the i^{th} stockout occurs, i.e., $\sum_{k=1}^{\kappa_i}\tau^k = t_i$. For all products, set $z^k_j = 1$ for $k \le \kappa_1$. From (18), we know that $a^k_j = \alpha_j(\mathbf{1})$ for all j and $k \le \kappa_1$. Moreover, for product $j(1)$, the first product to stock out, set $z^k_{j(1)} = 0$ for $k > \kappa_1$, consistent with (8). From (18a), we then have $a^k_{j(1)} = 0$ for $k > \kappa_1$. Note that (9) requires $y_{j(1)} = \sum_{k=1}^{\kappa_1}\lambda\alpha_{j(1)}(\mathbf{1})\tau^k = \lambda\alpha_{j(1)}(\mathbf{1})t_1 = y^F_{j(1)}$.

Now, let us consider some product $j(i) \in \mathbb{J}^-$ that stocks out at t_i. First, set $z^k_j = 1$ for all products that stock out at or after t_i, i.e., $j \in \{1, 2, \ldots, n\} \setminus \{j(1), j(2), \ldots, j(i-1)\}$, and $\kappa_{i-1} < k \le \kappa_i$. For any $\kappa_{i-1} < k \le \kappa_i$, we have $a_j^k = \alpha_j(\mathbf{1} - \sum_{i'=1}^{i-1} \mathbf{e}_{j(i')})$ due to (18). In addition, setting $z_{j(i)}^k = 0$ for all $k > \kappa_i$ requires $a_{j(i)}^k = 0$ for all $k > \kappa_i$ from (18a). Thus, computing (9) yields $y_{j(i)} = \sum_{k=1}^{\kappa_1} \lambda\alpha_{j(1)}(\mathbf{1})\tau^k + \sum_{i'=2}^{i}\sum_{k=\kappa_{i'-1}}^{\kappa_{i'}} \lambda\alpha_{j(i)}(\mathbf{1} - \sum_{\ell=1}^{i'-1} \mathbf{e}_{j(\ell)})\tau^k = y_{j(i)}^F$. Progressing in this manner, and consistent with (8), we can establish the following values for

$$z_{j(i)}^k = 1, \qquad \forall i = 1, 2, \ldots, \eta(\mathbf{y}^F);\ k \le \kappa_i$$
$$z_{j(i)}^k = 0, \qquad \forall i = 1, 2, \ldots, \eta(\mathbf{y}^F);\ k > \kappa_i$$

Notice that substituting these values in (18) ensures that, for all $j = 1, 2, \ldots, n$, we have

$$\begin{aligned}
a_j^k &= \alpha_j(\mathbf{1}), && k \le \kappa_1\\
a_j^k &= \alpha_j(\mathbf{1} - \mathbf{e}_{j(1)}), && \kappa_1 < k \le \kappa_2\\
&\vdots\\
a_j^k &= \alpha_j\Big(\mathbf{1} - \sum_{i=1}^{\eta(\mathbf{y}^F)-1} \mathbf{e}_{j(i)}\Big), && \kappa_{\eta(\mathbf{y}^F)-1} < k \le \kappa_{\eta(\mathbf{y}^F)}\\
a_j^k &= \alpha_j\Big(\mathbf{1} - \sum_{i=1}^{\eta(\mathbf{y}^F)} \mathbf{e}_{j(i)}\Big), && k > \kappa_{\eta(\mathbf{y}^F)}
\end{aligned}$$

With this $(\mathbf{y}^F, \mathbf{z}, \mathbf{a})$, for product $j(i) \in \mathbb{J}^-$, we evaluate (10) as

$$\sum_{k=1}^{m} z_{j(i)}^k \frac{\tau^k}{T} = \sum_{k=1}^{\kappa_i} \frac{\tau^k}{T} = \frac{t_{j(i)}}{T} = \tilde{\omega}_{j(i),T}(\mathbf{y}^F) \tag{34}$$

For a product $j \in \mathbb{J}^+$, we have $z_j^k = 1$ for all $k = 1, 2, \ldots, m$. Therefore,

$$\sum_{k=1}^{m} z_j^k \frac{\tau^k}{T} = \sum_{k=1}^{m} \frac{\tau^k}{T} = \frac{T}{T} = 1 = \tilde{\omega}_{jT}(\mathbf{y}^F) \tag{35}$$

Next, consider the category ready rate corresponding to $(\mathbf{y}^F, \mathbf{z}, \mathbf{a})$, as specified in (11).

$$\begin{aligned}
\sum_{k=1}^{m} \frac{\sum_{j=1}^n a_j^k}{\sum_{j=1}^n \alpha_j(\mathbf{1})}\frac{\tau^k}{T} &= \sum_{k=1}^{\kappa_1} \frac{\sum_{j=1}^n \alpha_j(\mathbf{1})}{\sum_{j=1}^n \alpha_j(\mathbf{1})}\frac{\tau^k}{T} + \sum_{k=\kappa_1+1}^{\kappa_2} \frac{\sum_{j=1}^n \alpha_j(\mathbf{1} - \mathbf{e}_{j(1)})}{\sum_{j=1}^n \alpha_j(\mathbf{1})}\frac{\tau^k}{T}\\
&\quad + \ldots + \sum_{k=\kappa_{\eta(\mathbf{y}^F)-1}+1}^{\kappa_{\eta(\mathbf{y}^F)}} \frac{\sum_{j=1}^n \alpha_j(\mathbf{1} - \sum_{i=1}^{\eta(\mathbf{y}^F)-1} \mathbf{e}_{j(i)})}{\sum_{j=1}^n \alpha_j(\mathbf{1})}\frac{\tau^k}{T}\\
&\quad + \sum_{k=\kappa_{\eta(\mathbf{y}^F)}+1}^{m} \frac{\sum_{j=1}^n \alpha_j(\mathbf{1} - \sum_{i=1}^{\eta(\mathbf{y}^F)} \mathbf{e}_{j(i)})}{\sum_{j=1}^n \alpha_j(\mathbf{1})}\frac{\tau^k}{T}\\
&= \frac{\sum_{j=1}^n \alpha_j(\mathbf{1})}{\sum_{j=1}^n \alpha_j(\mathbf{1})}\frac{t_1}{T} + \frac{\sum_{j=1}^n \alpha_j(\mathbf{1} - \mathbf{e}_{j(1)})}{\sum_{j=1}^n \alpha_j(\mathbf{1})}\frac{(t_2 - t_1)}{T}\\
&\quad + \ldots + \frac{\sum_{j=1}^n \alpha_j(\mathbf{1} - \sum_{i=1}^{\eta(\mathbf{y}^F)-1} \mathbf{e}_{j(i)})}{\sum_{j=1}^n \alpha_j(\mathbf{1})}\frac{(t_{\eta(\mathbf{y}^F)} - t_{\eta(\mathbf{y}^F)-1})}{T}\\
&\quad + \frac{\sum_{j=1}^n \alpha_j(\mathbf{1} - \sum_{i=1}^{\eta(\mathbf{y}^F)} \mathbf{e}_{j(i)})}{\sum_{j=1}^n \alpha_j(\mathbf{1})}\frac{(T - t_{\eta(\mathbf{y}^F)})}{T}\\
&= \tilde{\Omega}_T(\mathbf{y}^F)
\end{aligned} \tag{36}$$

If (34), (35) and (36) evaluate the ready rates that are at least as much as their corresponding targets, then $(\mathbf{y}^F, \mathbf{z}, \mathbf{a})$ is a feasible solution for [A-CIP^H].

Observe that the ready rate expressions in (31), (32), (33) for $\mathbf{y}^F$ and (34), (35), (36) for $(\mathbf{y}^F, \mathbf{z}, \mathbf{a})$ are indeed equivalent. As a result, this implies that $\mathbf{y}^F$ is feasible if and only if $(\mathbf{y}^F, \mathbf{z}, \mathbf{a})$. □

A.2. Proof of Proposition 5

Consider the process $\mathbf{X}(s) = \mathbf{y} - \mathfrak{N}(s)$, where $\mathfrak{N}(s)$ is the Poisson process of customer arrivals with rate $\lambda\alpha(\mathbf{X}(s))$. Suppose the firm runs out of product j at $T_j \le T$. Any feasible solution must satisfy:

$$y_j \geq \int_0^{T_j} \lambda\alpha_j(\mathbf{X}(s))ds \geq T_j\lambda\alpha_j(\mathbf{1}) \tag{37}$$

Naturally, the optimal solution y_j^* must also satisfy (37). Accordingly, since $\omega_{jT}(y) = E[\frac{T_j}{T}|\mathbf{X}(0) = \mathbf{y}^*] \geq \phi_j$,

$$y_j^* = E\left[y_j^*|\mathbf{X}(0) = \mathbf{y}^*\right] \geq \lambda\alpha_j(\mathbf{1})E[T_j|\mathbf{X}(0) = \mathbf{y}^*] \geq \lambda\alpha_j(\mathbf{1})T\phi_j. \tag{38}$$

In turn, the optimal objective value must be bounded as follows:

$$\sum_{j=1}^{n} c_j y_j^* \geq \lambda T \sum_{j=1}^{n} c_j\alpha_j(\mathbf{1})\phi_j = \text{LB}_{[\text{CIP}]} \tag{39}$$

We next develop a feasible solution for [CIP], since such a solution would be an upper bound on the optimal objective function value. Let T_j be the stockout time of product j, i.e., $X_j(s) > 0$, if $s < T_j$ and $X_j = 0$ for $s \geq T_j$. When $s \leq T_j$, we have

$$E[\mathfrak{N}_j(s)] \leq \lambda s\alpha_j(\mathbf{e}_j) \rightarrow E[X_j(s)|\mathbf{X}(0) = \mathbf{y}] \geq y_j - \lambda\alpha_j(\mathbf{e}_j)s \tag{40}$$

$$\rightarrow 0 = E[X_j(T_j)|\mathbf{X}(0) = \mathbf{y}] \geq y_j - \lambda\alpha_j(\mathbf{e}_j)T_j \tag{41}$$

$$\rightarrow E[T_j|\mathbf{X}(0) = \mathbf{y}] \geq \frac{y_j}{\lambda\alpha_j(\mathbf{e}_j)} \tag{42}$$

From (1) in the paper, we know that $\omega_{jT}(\mathbf{y}) = E\left[\frac{1}{T}\int_0^T \pi_j(\mathbf{X}(s))ds|\mathbf{X}(0) = \mathbf{y}\right] = E[\frac{T_j}{T}|\mathbf{X}(0) = \mathbf{y}] \geq \frac{y_j}{\lambda\alpha_j(\mathbf{e}_j)T}$. That is, if we set $y_j \geq \lambda\alpha_j(\mathbf{e}_j)\phi_j T$, we can show that $\omega_{jT}(\mathbf{y}) \geq \phi_j$. Now, consider the category ready rate constraint (3) in the paper,

$$\begin{aligned}\Omega_t &= E\left[\frac{1}{T}\int_0^T \Pi(\mathbf{X}(s))ds|\mathbf{X}(0) = \mathbf{y}\right] \\ &= \frac{1}{\sum_{j=1}^n \alpha_j(\mathbf{1})}\frac{1}{T}\sum_{j=1}^{n} E\left[\int_0^T \alpha_j(\mathbf{X}(s))ds|\mathbf{X}(0) = \mathbf{y}\right]\end{aligned} \tag{43}$$

Therefore,

$$\begin{aligned}E\left[\int_0^T \alpha_j(\mathbf{X}(s))ds|\mathbf{X}(0) = \mathbf{y}\right] &= E\left[\int_0^{T_j} \alpha_j(\mathbf{X}(s))ds|\mathbf{X}(0) = \mathbf{y}\right] \\ &\geq E\left[\int_0^{T_j} \alpha_j(\mathbf{1})ds|\mathbf{X}(0) = \mathbf{y}\right] \\ &= \alpha_j(\mathbf{1})E\left[T_j|\mathbf{X}(0) = \mathbf{y}\right] \geq \frac{y_j\alpha_j(\mathbf{1})}{\lambda\alpha_j(\mathbf{e}_j)},\ 1 \leq j \leq n.\end{aligned} \tag{44}$$

and consequently from (43) and (44),

$$\Omega_t(\mathbf{y}) \geq \frac{1}{\sum_{j=1}^n \alpha_j(1)}\frac{1}{T}\sum_{j=1}^{n}\frac{y_j\alpha_j(\mathbf{1})}{\lambda\alpha_j(\mathbf{e}_j)}. \tag{45}$$

Suppose β_j is a positive weight factor such that $\sum_j \beta_j = 1$. Then, if

$$y_j \geq \frac{\beta_j \alpha_j(\mathbf{e}_j) T \Phi \lambda \sum_{j=1}^n \alpha_j(\mathbf{1})}{\alpha_j(\mathbf{1})} \rightarrow \Omega_t(\mathbf{y}) \geq \frac{1}{\sum_{j=1}^n \alpha_j(\mathbf{1})} \frac{1}{T} \sum_{j=1}^n \frac{y_j \alpha_j(\mathbf{1})}{\lambda \alpha_j(\mathbf{e}_j)} \geq \Phi, \tag{46}$$

Let $\beta_j = 1/n$. Therefore,

$$y_j = \max\left[\lceil \frac{\alpha_j(\mathbf{e}_j) T \Phi \lambda \sum_{j=1}^n \alpha_j(\mathbf{1})}{n \alpha_j(\mathbf{1})} \rceil, \lceil \lambda \alpha_j(\mathbf{e}_j) \phi_j T \rceil\right] \tag{47}$$

satisfies both ready rate constraints and is feasible for [CIP] and its total cost is

$$\text{UB}_{[\text{CIP}]} = \mathbf{c}.\mathbf{y} = \sum_{j=1}^n \max\left[c_j \lceil \frac{\alpha_j(\mathbf{e}_j) T \Phi \lambda \sum_{j=1}^n \alpha_j(\mathbf{1})}{n \alpha_j(\mathbf{1})} \rceil, c_j \lceil \lambda \alpha_j(\mathbf{e}_j) \phi_j T \rceil\right] \tag{48}$$

□

A.3. Proof of Proposition 6

To satisfy the product ready rate constraints ((10) in [A-CIP]), product j must be available for at least $\bar{k}_j = \min[K | \sum_{k=1}^K \tau^k \geq \phi_j T]$ time intervals. Therefore, the firm must carry at least $\sum_{k=1}^{\bar{k}_j} \lambda \tau^k \alpha_j(\mathbf{1})$ units of inventory of product j. Since $\alpha_j(\mathbf{1}) \leq a_j^k$ for any $1 \leq k \leq \bar{k}_j$, the optimal solution $\tilde{y}_j$ satisfies

$$\tilde{y} \;\geq\; \sum_{k=1}^m \lambda \tau^k a_j^k \;\geq\; \sum_{k=1}^{\bar{k}_j} \lambda \tau^k a_j^k \;\geq\; \sum_{k=1}^{\bar{k}_j} \lambda \tau^k \alpha_j(\mathbf{1}) \tag{49}$$

The corresponding lower bound on the optimal objective function value is

$$\sum_{j=1}^n c_j \tilde{y}_j \;\geq\; \sum_{j=1}^n c_j \sum_{k=1}^{\bar{k}_j} \lambda \tau^k \alpha_j(\mathbf{1}) \;\geq\; \min_{1 \leq j \leq n} \alpha_j(\mathbf{1}) \sum_{j=1}^n c_j \sum_{k=1}^{\bar{k}_j} \lambda \tau^k \;\geq\; \lambda T \min_{1 \leq j \leq n} \alpha_j(\mathbf{1}) \sum_{j=1}^n c_j \phi_j \tag{50}$$

We can also obtain another lower bound as follows:

$$\sum_{j=1}^n c_j \tilde{y}_j \;\geq\; \sum_{j=1}^n c_j \sum_{k=1}^m \lambda \tau^k a_j^k \;\geq\; \lambda \min_{1 \leq j \leq n} \{c_j\} \sum_{j=1}^n \sum_{k=1}^m \tau^k a_j^k \;\geq\; \lambda T \Phi \min_{1 \leq j \leq n} \{c_j\} \sum_{j=1}^n \alpha_j(\mathbf{1}), \tag{51}$$

where the last step stems from the category ready rate constraint, i.e.,

$$\sum_{k=1}^m \left(\frac{\sum_{j=1}^n a_j^k}{\sum_{j=1}^n \alpha_j(\mathbf{1})} \cdot \frac{\tau^k}{T} \right) \geq \Phi \longrightarrow \sum_{k=1}^m \sum_{j=1}^n a_j^k \cdot \tau^k \geq T \Phi \sum_{j=1}^n \alpha_j(\mathbf{1}). \tag{52}$$

Neither lower bound dominates the other uniformly, so we take the maximum of the two as the lower bound. That is,

$$\text{LB}_{[\text{A-CIP}]} = \max\left[\lambda T \min_{1 \leq j \leq n} \alpha_j(\mathbf{1}) \sum_{j=1}^n c_j \phi_j, \lambda T \Phi \min_{1 \leq j \leq n} \{c_j\} \sum_{j=1}^n \alpha_j(\mathbf{1})\right] \tag{53}$$

To construct an upper bound on the optimal value of [A-CIP], consider a natural feasible solution that ensures that all the products are available during the entire planning horizon T. This solution sets:

$$\begin{cases} z_j^k = 1, \; \forall 1 \leq k \leq m \\ y_j = \lceil T \lambda \alpha_j(\mathbf{1}) \rceil, \; \forall 1 \leq j \leq n \\ a_j^k = \alpha_j(\mathbf{1}), \; \forall 1 \leq k \leq m \end{cases} \tag{54}$$

Constraints (8), (9), (13) - (15) in [A-CIP] follow directly from (54). Additionally, for (10) and (11) in [A-CIP], we have:

$$\sum_{k=1}^m \left(z_j^k \frac{\tau^k}{T} \right) = 1 \geq \phi_j, \text{ and} \tag{55}$$

$$\sum_{k=1}^{m}\left(\frac{\sum_{j=1}^{n} a_j^k}{\sum_{j=1}^{n}\alpha_j(\mathbf{1})}\cdot\frac{\tau^k}{T}\right)=\sum_{j=1}^{n}\left(\frac{T\alpha_j(\mathbf{1})}{\sum_{j=1}^{n}T\alpha_j(\mathbf{1})}\right)=1\geq\Phi. \tag{56}$$

Therefore, this feasible solution yields the following upper bound on the optimal objective function value of [A-CIP]:

$$\mathrm{UB}_{[\text{A-CIP}]}=\sum_{j=1}^{n}c_j\lceil T\lambda\alpha_j(\mathbf{1})\rceil \tag{57}$$

□

B. Procedure for Random Problem Generation

We follow the following procedure to generate random instances of the CIP problem.

(1) Select number of products n as per computational experiment design.

(2) Select market share structure (*equal* versus *unequal*).

(3) Let $\boldsymbol{\alpha}$ denote the market share vector (with full availability of all products in category).

(a) For equal market shares, set $\boldsymbol{\alpha}=\{\frac{0.99}{n},\frac{0.99}{n},\cdots,\frac{0.99}{n}\}$ and the outside option share as 0.01.

(b) For unequal market shares, set $\boldsymbol{\alpha}=\left\{\frac{0.99}{n(n+1)/2},2\times\frac{0.99}{n(n+1)/2},3\times\frac{0.99}{n(n+1)/2},\cdots,n\times\frac{0.99}{n(n+1)/2}\right\}$ and the outside option share as 0.01.

(4) Select demand rate $\lambda\in\{25n,50n,100n\}$, which reflect low, moderate and high levels of demand, respectively, for each product.

(5) Select category ready rate $\Phi\in\{85\%,90\%,95\%\}$.

(6) Select product ready rate $\phi_j\in\{0.90\times\Phi,0.95\times\Phi,0.99\times\Phi\}$ for $j=\{1,2,\cdots,n\}$.

(7) Select choice model as either HORIZONTAL or VERTICAL.

(a) If the choice model is HORIZONTAL:

i. Normalize $\beta_n=100$. Next, for every $i<n$, set $\beta_i=\beta_{i+1}\times(1+\zeta)$ where ζ is a random draw from the uniform distribution between 0 and 0.1.

ii. Note that the product price needs to be set as consistent with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to reflect the nature of the MNL (horizontal) choice model. Accordingly, we normalize the value of the outside option to zero, and let $\epsilon=1/(1+\sum_{j=1}^{n}e^{\beta_j-p_j})$, and $\alpha_i=\frac{e^{\beta_i-p_i}}{1+\sum_{j=1}^{n}e^{\beta_j-p_j}}$. Consequently, we have $\frac{\alpha_i}{\alpha_n}=e^{\beta_i-p_i}$. Equivalently, $\ln\alpha_i-\ln\alpha_n=\beta_i-p_i$, and $p_i=\beta_i+\ln\epsilon-\ln\alpha_i$ for all i.

iii. Set $c_i=100$ for all i.

(b) If the choice model is VERTICAL:

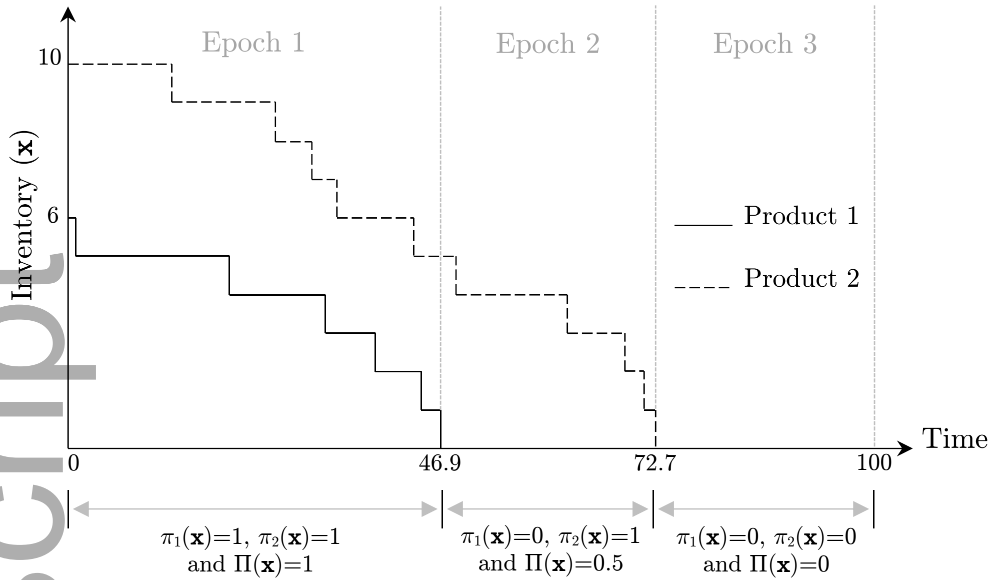
i. Normalize $q_n=1$. Next, for every $i<n$, set $q_i=\lceil q_{i+1}\times(1+\zeta)\rceil$ where ζ is a random draw from the uniform distribution between 0 and 1.

ii. Note that the product price needs to be set as consistent with $\boldsymbol{\alpha}$ and $\mathbf{q}$ to reflect the nature of the vertical choice model. Accordingly, we need to solve for $\mathbf{p}$ in the following set of equations:

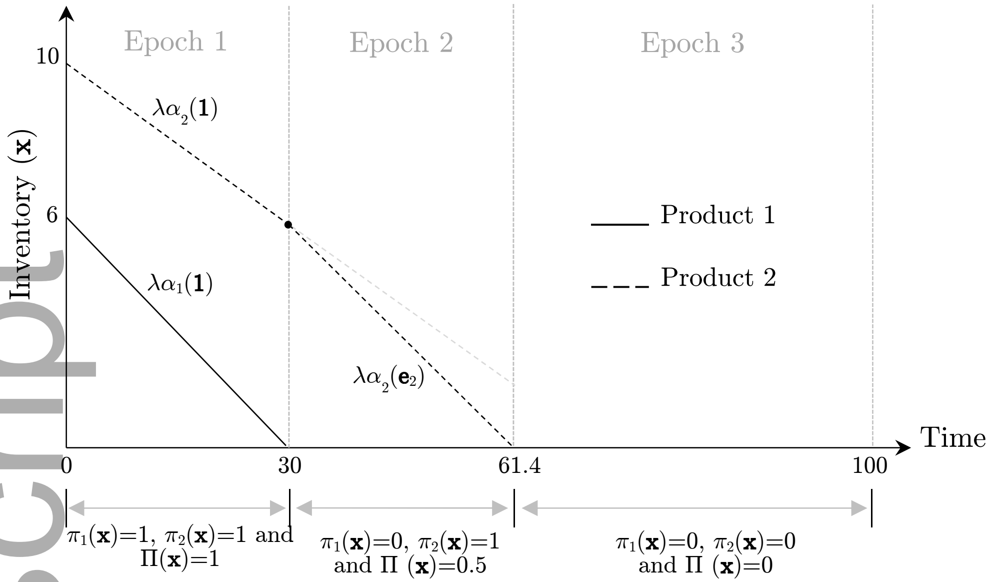
$$\begin{aligned}1-\frac{p_1-p_2}{q_1-q_2}&=\alpha_1\\ \frac{p_{i-1}-p_i}{q_{i-1}-q_i}-\frac{p_i-p_{i+1}}{q_i-q_{i+1}}&=\alpha_i\quad i\in\{2,3,\ldots,n-1\}\\ \frac{p_{n-1}-p_n}{q_{n-1}-q_n}-\frac{p_n}{q_n}&=\alpha_n\end{aligned}$$

We first set $p_n=q_n(1-\sum_{i=1}^{n}\alpha_i)$, and solve for p_{n-1} in $\frac{p_{n-1}-p_n}{q_{n-1}-q_n}-\frac{p_n}{q_n}=\alpha_n$, and so on.

iii. Set $c_n = 100$ and $c_i = c_{i+1}(1+\varrho)$ where ϱ is a random draw from the uniform distribution between 0 and 1.



poms_13240_f1.eps



poms_13240_f2.eps